Confounding Dynamics*

Giacomo Rondina[†] Todd B. Walker[‡]

January 2017

Abstract

In the context of a dynamic model with incomplete information, we isolate a novel mechanism of shock propagation that results in waves of optimism and pessimism along a Rational Expectations equilibrium. We term the mechanism *confounding dynamics* because it arises from agents' optimal signal extraction efforts on variables whose dynamics—as opposed to superimposed noise—prevents full revelation of information. Employing methods in the space of analytic functions, we are able to obtain analytical characterizations of the equilibria that generalize the celebrated Hansen-Sargent optimal prediction formula. We apply our results to a canonical real business cycle model and derive the analytic solution for output, consumption and capital. We show that, in response to a permanent positive productivity shock, confounding dynamics generate expansions *and* recessions that would not be present under complete information.

Keywords: Dispersed Information, Confounding Dynamics, Rational Expectations

*We would like to thank Dimitri Vayanos and three anonymous referees for very useful comments. We would also like to thank Manuel Amador, Marios Angeletos, Ryan Chahrour, Tim Cogley, Thorsten Drautzburg, Jesús Fernández-Villaverde, Jennifer La'O, Eric Leeper, Thomas Mertens, Kristoffer Nimark, Gary Ramey, Tom Sargent, Martin Schneider, Venky Venkateswaran, Pierre-Olivier Weill, Mirko Wiederholt, Chuck Whiteman and seminar participants at UC–Berkeley, UC–Santa Cruz, New York University, University of Pennsylvania, University of Texas–Austin, Stanford University, St. Louis Federal Reserve, the Econometric Society 2009 summer meetings, the 2009 Yale University Cowles Foundation Summer Conference on Information and Beliefs in Macroeconomics, the 2011 LAEF Spring Conference on Information in Macroeconomics, the 2011 CREI conference on Beliefs and Expectations in Macroeconomics, and the 2012 ASSA meetings for useful comments. We also acknowledge financial support from the National Science Foundation under grant number SES–0962221. Rondina would also like to thank the Rady School of Management at UCSD for the kind hospitality.

[†]University of Colorado at Boulder, giacomo.rondina@colorado.edu

[‡]Indiana University, walkertb@indiana.edu

1 Introduction

Modeling and seeking to understand economic fluctuations is one of the cornerstones of modern economics. The role of incomplete information in this endeavor was acknowledged very early on by Pigou (1929) and Keynes (1936). Their ideas were first formalized in a rational expectations setting by Lucas (1972, 1975), King (1982) and Townsend (1983b). The underlying theme that ties these papers together is that unresolved uncertainty—in and of itself—can be a source of fluctuation in the economy. This idea has seen a resurgence. Dynamic models with dispersed information are becoming increasingly prominent in several literatures such as asset pricing, optimal policy communication, international finance, and business cycles.[1] Our paper contributes to this literature by introducing the concept of confounding dynamics and doing so in a manner that permits tractability.

Confounding dynamics arise from optimal prediction (i.e. rational expectations) in which past realizations of economic shocks prevent full revelation of information today, even when an arbitrarily large amount of data is available. Ensuring confounding dynamics survive amounts to deriving (non-invertibility) restrictions on equilibrium dynamics. If endogenous variables are non-invertible in current and past observations, agents will never fully unravel the contemporaneous economic shock. This is true even when the number of shocks is equal to the number of observables. Thus, a distinguishing feature of our approach is that confounding dynamics does not rely on the need to impose exogenous noise on the model.[2]

The defining property of confounding dynamics is that the equilibrium impulse response functions consist of oscillating over- and under-reactions relative to the full information (or exogenous noise) equilibrium. The typical shape of an impulse response to a fundamental shock under complete information is one that monotonically decays back to steady state either from above or below. With confounding dynamics, impulse response functions fluctuate around the full information counterpart. That is, they display the "waves of optimism and pessimism" of Pigou (1929).

Section 2.2 derives an optimal prediction formula under confounding dynamics that extends the celebrated Hansen-Sargent formula, and makes an explicit connection to these dynamics. Section 4.2 demonstrates that this behavior carries over to a generic rational expectations model with dispersed information. Our main theorem contains two equations—one that characterizes the dynamic properties of the equilibrium when confounding dynamics are present and one that derives exogenous restrictions that guarantee confounding dynamics are preserved in equilibrium. Section 5

[1] The literature is too voluminous to cite every worthy paper. Recent examples include: Woodford (2003a), Pearlman and Sargent (2005), Allen, Morris, and Shin (2006), Bacchetta and van Wincoop (2006), Hellwig (2006), Adam (2007), Gregoir and Weill (2007), Angeletos and Pavan (2007), Kasa, Walker, and Whiteman (2014), Lorenzoni (2009), Rondina (2009), Angeletos and La'O (2009), Angeletos and La'O (2013), Hellwig and Venkateswaran (2009), Graham and Wright (2010), Nimark (2010), Hassan and Mertens (2011), Benhabib, Wang, and Wen (2015), Huo and Takayama (2016).

[2] Many of the papers in the dynamic dispersed information literature introduce exogenous noise in order to preserve asymmetric information in equilibrium. While our approach can certainly accommodate exogenous noise, confounding dynamics can lead to dispersed information without having more shocks than signals.

provides intuition by introducing confounding dynamics into a standard Real Business Cycle model. To the best of our knowledge, this equilibrium dynamic behavior and formal characterization is new to the rational expectations literature.

We solve and analyze the rational expectations equilibrium in the space of analytic functions. This approach has several advantages vis-a-vis standard time-domain methods. For example, as emphasized in Townsend (1983a), equilibria are sought in generic functional spaces spanned by linear combinations of shocks, which allows one to avoid explicitly modeling higher-order belief dynamics (Section 3.5). Moreover, the matrix Ricatti equation is replaced by a more transparent spectral factorization problem. This allows us to solve and analyze the equilibrium in closed form.

We are not the first to advocate such an approach. Others, such as Futia (1981), Townsend (1983a), Taub (1989), Kasa (2000), Walker (2007), Rondina (2009), Bernhardt, Seiler, and Taub (2010), Kasa, Walker, and Whiteman (2014), and Huo and Takayama (2016) have used similar techniques to solve dynamic rational expectation models with incomplete information. We contribute to this literature by deriving analytical representations (e.g., generalized Hansen-Sargent formulas) and by providing a systematic treatment of equilibrium conditions in models with *dispersed* information that display confounding dynamics. Futia (1981) and Townsend (1983a) were the first to advocate for the use of analytic functions to solve dynamic rational expectations models with heterogeneous information. Many of the mathematical antecedents of this paper can be found there and in Whiteman (1983). Taub (1989) demonstrates how the algebra associated with dynamic signal extraction (i.e., spectral factorization) is simplified through the analytic function approach. We take advantage of these formulas to completely characterize existence and uniqueness of equilibria in dispersed informational setups. Bernhardt, Seiler, and Taub (2010) and Kasa, Walker, and Whiteman (2014) do not examine models with dispersed information, but show how these methods can be used to help resolve asset pricing anomalies.

2 Preliminaries

In this section, we first establish notation and introduce relevant mathematical definitions. We then formalize the notion of confounding dynamics, presenting a simple example that shows the mechanism at work.

2.1 Mathematical Preliminaries

Throughout the paper, we work in the space of polynomials in the lag operator L with square-summable coefficients that operate on Gaussian random variables. In our framework, any stochastic process ω_t can always be written as

$$\omega_t = Q(L)\varepsilon_t = \sum_{j=0}^{\infty} Q_j L^j \varepsilon_t, \tag{2.1}$$

where $\sum_{j=0}^{\infty} |Q_j|^2 < \infty$, and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$, are innovations identically and independently distributed over time. In linear-Gaussian environments, working with representations of the form of (2.1), and their functional equivalents, has three advantages for analyzing rational expectations models with incomplete information.[3]

First, representation (2.1) is general in the sense that it can accommodate both auto-regressive (AR) and moving-average (MA) components of any order. This is especially useful when searching for an equilibrium because it avoids the need to specify a conjecture with a specific ARMA order. Regardless of the complexity of the equilibrium conditions that emerge in models of dispersed information (e.g., infinite regress in expectations), the solution will take the form of (2.1).[4]

Second, the Wold Representation Theorem ensures that processes like ω_t can always be written *uniquely* as a linear combination of a moving average representation where the innovations are the linear forecast errors for ω_t, conditional on any linear-Gaussian information set [Brockwell and Davis (1987)]. That is, the Wold representation establishes the invertibility of $Q(L)$ and one may write $Q(L)^{-1}\omega_t = \varepsilon_t$, which implies that the space spanned by $\{\omega_t, \omega_{t-1}, ...\}$ is equivalent (in mean-square norm) to the space spanned by $\{\varepsilon_t, \varepsilon_{t-1}, ...\}$. Consequently, one can apply the optimal prediction formulas derived by Wiener-Kolmogorov [Whittle (1983)] to compute the conditional expectation of processes like ω_t.

Third, the Riesz-Fischer Theorem [see Sargent (1987)] establishes an isometric, isomorphic mapping from the space of lag polynomials with square-summable coefficients $Q(L)$ to the space of analytic complex-valued functions, where (2.1) is represented as $Q(z)$, but with $z \in \mathbb{C}$. In several key steps of the analysis in this paper we find it convenient to exploit the properties of such functions, which allows us to derive simple existence and uniqueness conditions for rational expectations equilibria with incomplete and dispersed information following Whiteman (1983). In a slight abuse of notation, we employ L and z interchangeably when working in the space of analytic functions.

While this methodology is extremely helpful in solving dynamic models with incomplete information, it is not well known by economists. Thus, we provide an Online Technical Appendix with the statements of the key theorems cited above and further references for interested readers. We now restrict our focus to the formulation of optimal prediction formulas, which is where confounding dynamics emerge.

Suppose that we would like to formulate the prediction of ω_{t+j} so as to minimize the mean-squared forecast error, conditional on the observation of the history of a $n \times 1$ vector of variables, s_t, up to time t. To denote such history, we use the compact notation, $s^t \equiv \{s_{t-j}\}_{j=0}^{\infty}$. Let

$$s_t = \Gamma(L)u_t, \tag{2.2}$$

where $\Gamma(L)$ is an $n \times m$ matrix with each element being a square-summable lag polynomial in non-negative powers of L, and u_t an $n \times 1$ vector of i.i.d. Gaussian shocks with variance-covariance normalized to the $n \times n$ identity matrix. Let $g_{ss}(z)$ be the variance-covariance gener-

[3] We are not the first to highlight these advantages [see, Hansen and Sargent (1980), Futia (1981), Townsend (1983b), Kasa (2000)].

[4] Townsend (1983a) elaborates extensively on the advantages of using representations such as (2.1) when solving for equilibria that harbor an infinite regress in expectations. In Section 3.5 we show the complex form that the infinite regress in expectations takes in our framework.

ating function for the process s_t,[5] then one has that $g_{ss}(z) = \Gamma(z)\Gamma(z^{-1})^\top$. Similarly, one can define the covariance generating function between the joint processes ω_t and s_t, which is given by $g_{\omega s}(z) = Q(z)\sigma_\varepsilon\Gamma(z^{-1})^\top$. The prediction for ω_{t+j} that minimizes the mean squared forecast error corresponds to a linear combination of current and past realizations of s_t, denoted by

$$\mathcal{P}(\omega_{t+j}|s^t) = \Pi(L)s_t \tag{2.3}$$

Here, $\Pi(L)$ is a $1 \times n$ vector of square-summable lag polynomials in non-negative powers of L, whose form is provided by the Wiener-Kolmogorov formula[6]

$$\Pi(L) = \left[L^{-j}g_{\omega s}(L)\left(\Gamma^*(L^{-1})^\top\right)^{-1}\right]_+\Gamma^*(L)^{-1} \tag{2.4}$$

Expression (2.4) has several moving parts that require some unpacking. Let us first consider the special case $s_t = \omega_t$, which implies $\Gamma(L) = Q(L)$, so that the prediction problem is one in which we would like to predict the future realizations ω_{t+j}, for $j \geq 1$, using its own past, ω^t. Given the form of ω_t in (2.1), the best prediction is one that carries $Q(L)$ j periods forward and uses the best estimates of the infinite history of innovations $\{\varepsilon_{t+j}, \varepsilon_{t+j-1}, ...\}$, to compute ω_{t+j}. The best estimates of $\{\varepsilon_{t+j}, \varepsilon_{t+j-1}, ..., \varepsilon_{t+1}\}$ are clearly equal to the unconditional average, 0. It follows that they should not appear in the optimal prediction. This is achieved by the operator $[\ \ ]_+$ in (2.4), known as the "annihilating operator", which instructs us to ignore the first j coefficients of its argument.

The best estimates for $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, on the other hand, are usually not zero, and correspond to the innovations in the Wold fundamental representation for ω_t. The decomposition of the information set into its Wold fundamental representation is achieved by $\Gamma^*(L)$ in (2.4). Whether $\Gamma^*(L)$ is equal to $\Gamma(L)$ depends on the invertibility properties of the analytic function $\Gamma(z)$. For our special case, if $Q(z)$ is invertible for z inside the unit circle, then the Wold fundamental representation corresponds exactly with the history $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, and the prediction formula $\mathcal{P}(\omega_{t+j}|\omega^t) = \tilde{\Pi}(L)\omega_t$, has

$$\tilde{\Pi}(L) = [L^{-j}Q(L)]_+Q(L)^{-1} \tag{2.5}$$

Note that all the steps just described are clearly at work here: L^{-j} carries the function $Q(L)$ j periods forward; the operator $[\ \ ]_+$ sets the estimates of innovations from $t+1$ to $t+j$ to zero, by annihilating Q_0, Q_1,...,Q_{j-1}; finally, $Q(L)^{-1}$ makes sure that once $\Pi(L)$ is multiplied by ω_t, the Wold fundamental innovations $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$ result.

Expression (2.5) is a special case of (2.4) because, in general, $\Gamma(z)$ is not invertible for z inside the unit circle, and solving for $\Gamma^*(z)$ is at the core of the solution to the prediction problem. Formally, $\Gamma^*(z)$ corresponds to the "canonical" factorization of the variance-covariance generating

[5] The variance-covariance generating function of a stationary Gaussian process is defined as the Fourier transform of its correlogram, which is the collection of covariances at all horizons. See the Online Technical Appendix for details.

[6] See Whittle (1983), or the Online Technical Appendix, for a derivation of the formula.

function $g_{ss}(z)$, so that

$$g_{ss}(z) = \Gamma^*(z)\Gamma^*(z^{-1})^\top \tag{2.6}$$

The canonical factorization answers the question: What space is spanned by the observables (s^t)? If $\Gamma(z)$ is invertible in non-negative powers of z, then s^t will span the innovations u^t. If $\Gamma(z)$ is non-invertible, then $\Gamma^*(z)$ must be determined and it will span a space that is strictly smaller than that spanned by u^t. The existence of $\Gamma^*(z)$ is guaranteed by the Wold (fundamental) Representation Theorem. The computation of $\Gamma^*(z)$ can be quite involved, however, and there exist several methods for achieving it. In our setting, we follow the steps outlined in Rozanov (1967).[7]

There could be two reasons for $\Gamma(z)$ to be non-invertible: (i) $\Gamma(z)$ could be "rectangular", so that $m > n$, which means that the dimension of the vector of underlying shocks, u_t, is greater than the vector of signals, s_t, at any t; (ii) $\Gamma(z)$ is square, so that $m = n$, and has one or more zeros inside the unit circle where its determinant vanishes. In case (i), non-invertibility is ensured by assumption as it is the direct consequence of assuming fewer signals than sources of noise. In case (ii), non-invertibility stems from the way signals combine over time and become themselves a source of noise, as the example in the next Section shows. Confounding dynamics can emerge in either scenario.

Definition CD. *If s_t is specified as in (2.2), then the s_t process is said to display confounding dynamics if there exists a λ, with $|\lambda| < 1$, such that* $\det \Gamma(z)|_{z=\lambda} = 0$.

If the determinant vanishes at $z = |\lambda| \in (-1, 1)$, then it guarantees that $\Gamma(z)$ is not invertible and therefore knowledge of s^t will *not* translate into direct knowledge of u^t $(u_t \neq \Gamma(L)^{-1}s_t)$. The Wold representation theorem implies that there always exists a $\Gamma(L)^*$ that is invertible. Deriving this $\Gamma(L)^*$ is the purpose of the next section.

The definition allows for an arbitrary number of roots at which the determinant of $\Gamma(L)$ vanishes. For simplicity, we will focus on characterizing equilibria with confounding dynamics due to only one root λ and for square systems, $m = n$. Appendix B.2 presents an example of confounding dynamics due to N non-invertible roots; while Appendix B.3 presents an example with a rectangular system and shows that the hallmark features of confounding dynamics are retained. In other words, the learning mechanism at the core of the equilibria that we characterize in Section 4 does not hinge upon the dimensions of $\Gamma(L)$ or the number of zeros inside the unit circle. Confounding dynamics

[7] Whittle (1983) shows that the computation of $\Gamma^*(L)$ is the dual to the solution of the n^{th} order Ricatti equation in the Kalman filter approach to optimal linear-quadratic prediction. In both cases the objective is to figure out the variance-covariance of the optimal prediction errors. Hence, *if* the Ricatti equation is solved, then $\Gamma^*(z)$ can be determined; and *if* $\Gamma^*(L)$ is computed, then a solution to the Ricatti equation can be determined. This is the approach taken in Huo and Takayama (2016). Whether it is easier to solve the Ricatti equation or to obtain the canonical factorization often depends on the application at hand, and whether one is looking for an analytical or a numerical solution. The canonical factorization, however, has a wider scope of applicability. Whittle (1983) presents a generalized approach – which he terms the Hamiltonian approach – where the prediction problem is solved by the canonical factorization of a matrix function obtained by augmenting the plant and observation equations of a Kalman system with the saddle point conditions of an Hamiltonian optimization problem. Huertgen, Hoffmann, Rondina, and Walker (2016) make an explicit connection between these solution methodologies.

can easily emerge when there are more shocks than signals ($m > n$) as shown in Appendix B.3.

2.2 Prediction with Confounding Dynamics To study the mechanism of confounding dynamics and showcase the usefulness of the Wiener-Kolmogorov prediction formula (2.4), we present a simple version of the prediction problem that operates at the heart of the rational expectations equilibria established in Section 4.

Let $Q(L) = 1$ in (2.1), so that $\omega_t = \varepsilon_t$, and s_t is a univariate process specified as

$$s_t = \lambda\varepsilon_t + \varepsilon_{t-1}, \tag{2.7}$$

which results in $\Gamma(L) = (\lambda + L)\sigma_\varepsilon$, $u_t = \sigma_\varepsilon^{-1}\varepsilon_t$, and we assume $\lambda > 0$. Suppose that the prediction problem is to compute the mean-squared error minimizing prediction for ε_t given that s^t is observed. To fix ideas, imagine that ε_t is the time-t unobserved innovation in aggregate productivity in the economy, while s_t is the observed market rental rate of physical capital. The prediction problem asks for an estimate of the current productivity innovation using the history of the market rental rate.

To solve the problem we can apply the Wiener-Kolmogorov prediction formula (2.4) with $j = 0$. Note that $g_{\omega s}(z) = \sigma_\varepsilon^2(\lambda + z^{-1})$, whereas to obtain $\Gamma^*(L)$, we need to consider two possible cases. If $\lambda \geq 1$, $\Gamma^*(L) = \Gamma(L) = (\lambda + L)\sigma_\varepsilon$, which means that the stochastic process (2.7) is invertible, and therefore there exists a linear combination of current and past s_t's that allows the exact recovery of ε_t. One can easily verify that applying (2.4) leads to $\Pi_0(L) = (\lambda + L)^{-1}$, and the optimal prediction corresponds to

$$\mathcal{P}(\varepsilon_t|s^t) = \frac{s_t}{\lambda + L} = 1/\lambda\big(s_t - \lambda^{-1}s_{t-1} + \lambda^{-2}s_{t-2} - \lambda^{-3}s_{t-3} + ...\big) = \varepsilon_t, \tag{2.8}$$

which verifies that the history of s^t contains all the information needed to perfectly know ε_t.

Consider now the case of $\lambda < 1$, so that s_t displays confounding dynamics, according to Definition CD. Clearly, the prediction formula (2.8) is no longer well defined as the coefficients diverge. The problem is that $\Gamma(L)$ is non-invertible at $L = -\lambda$ and so our previous choice for $\Gamma^*(L)$ would not work in this case. In this simple environment, Rozanov (1967) shows that the canonical factorization is given by flipping the root outside of the unit circle, $\Gamma^*(L) = (1 + \lambda L)\sigma_\varepsilon$, which results in

$$\Pi_0(L) = \Big[\sigma_\varepsilon^2(\lambda + L^{-1})\big[(1 + \lambda L^{-1})\sigma_\varepsilon\big]^{-1}\Big]_+ \big((1 + \lambda L)\sigma_\varepsilon\big)^{-1} = \lambda(1 + \lambda L)^{-1}. \tag{2.9}$$

The optimal prediction for ε_t is then

$$\mathcal{P}(\varepsilon_t|s^t) = \frac{\lambda}{1 + \lambda L}s_t = \lambda\big(s_t - \lambda s_{t-1} + \lambda^2 s_{t-2} - \lambda^3 s_{t-3} + ...\big) = \lambda\left(\frac{L + \lambda}{1 + \lambda L}\right)\varepsilon_t, \tag{2.10}$$

which results in a mean squared forecast error of $\left(1 - \lambda^2\right)\sigma_\varepsilon^2 > 0$, demonstrating that as λ approaches one from below there is exact recovery of ε_t.

When the process is non-invertible, (2.10) shows that the history of current and past s_t's reveals a particular linear combination of ε_t's. Expanding this last term yields

$$\mathcal{P}(\varepsilon_t|s^t) = \underbrace{\lambda^2\varepsilon_t}_{\text{information}} \quad + \quad \underbrace{(1-\lambda^2)[\lambda\varepsilon_{t-1} - \lambda^2\varepsilon_{t-2} + \lambda^3\varepsilon_{t-3} - \cdots]}_{\text{noise from confounding dynamics}} \tag{2.11}$$

Thus, the noise resulting from confounding dynamics takes an unusual form as it consists of a linear combination of past realizations of ε_t. Expression (2.11) suggests that the process (2.7) is informationally equivalent to a noisy signal about ε_t, where the noise is the linear combination of past shocks (in the bracketed term), and the signal-to-noise ratio is measured by λ^2. A λ closer to zero results in less information and more noise, but, at the same time, it also makes past shocks less persistent. In fact as $\lambda \to 0$, there is no information in s_t about ε_t and the optimal prediction is 0, the unconditional average. As long as $|\lambda| \in (-1, 1)$, the value of ε_t will *never* be learned and in this sense, the *history* of the fundamental shock acts as a standard noise shock. This is the defining characteristic of confounding dynamics.

To make the connection to the standard signal extraction problem more explicit, suppose that agents observe an infinite history of the signal

$$x_t = \varepsilon_t + \eta_t, \tag{2.12}$$

where $\eta_t \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right)$. The optimal prediction is well known and given by $\mathcal{P}(\varepsilon_t|x^t) = \tau x_t$, where τ is the relative weight given to the signal, $\tau = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$. We show in Appendix A that the information content of (2.7) with $|\lambda| < 1$ is equivalent to (2.12), where equivalence is defined as equality of variance of the forecast error conditioned on the infinite history of the observed signal, i.e.

$$\mathbb{E}\left[\left(\varepsilon_t - \mathcal{P}\left(\varepsilon_t|s^t\right)\right)^2\right] = \mathbb{E}\left[\left(\varepsilon_t - \mathcal{P}\left(\varepsilon_t|x^t\right)\right)^2\right],$$

when

$$\lambda^2 = \tau \tag{2.13}$$

Notice that when the signal-to-noise ratio increases (decreases), this corresponds to a higher (lower) absolute value of λ. In the limit, as $\sigma_\eta^2 \to 0$, then $\lambda^2 \to 1$, which ensures exact recovery of the state in both cases.

While the informational content can be made identical, the dynamics of the two signal extraction problems are very different. To visualize this, we report the impulse response function for the prediction equations that contain confounding dynamics (2.11) and for the standard signal extraction problem (2.12) to a one time, one unit increase in ε_t in Figure 1. We do this for both a low and high value of λ (τ).

First notice that with confounding dynamics, (2.11) under-predicts the actual innovation on impact, with a smaller value of λ under-predicting more significantly. This is due to the first term on the RHS of (2.11). The same is true for the standard signal extraction formulation (dashed lines). Agents weigh the initial innovation by the signal-to-noise ratio $\tau < 1$ and therefore

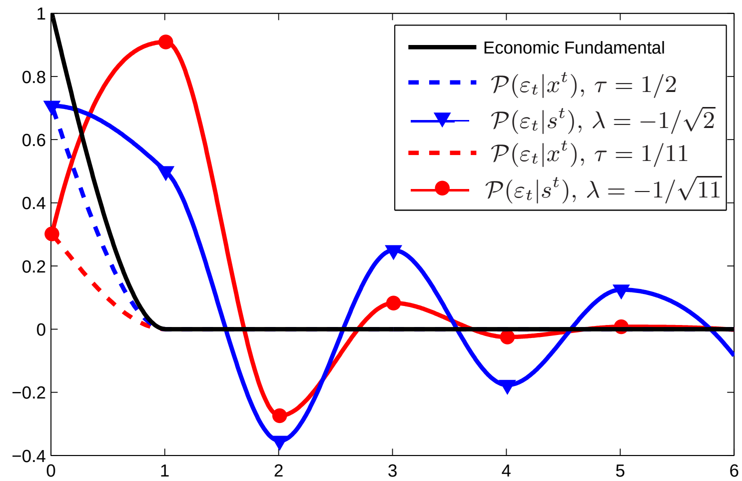


Figure 1: Impulse Responses of x_t and s_t to a one unit change in ε_t for signal-to-noise ratios of $\tau = 1/2$, $\lambda = -1/\sqrt{2}$ (dashed, triangle markers) and $\tau = 1/11$, $\lambda = -1/\sqrt{11}$ (dashed-dot, circle markers). The solid line is the innovation itself, ε_t, normalized to 1.

under-predict on impact. This is where the similarities end. With confounding dynamics, periods two through six show waves of over- and under-prediction relative to the actual realization and relative to the standard signal extraction problem. As discussed above, the current and past innovations will persistently affect the prediction function several periods beyond impact. This defining characteristic of confounding dynamics leads to the waves of over- and under-reaction. This is in contrast to the full information case and standard signal-extraction case where the impulse response is zero after impact. As already pointed out, the smaller the λ, the larger the noise term in (2.11), but the less persistent the over- and under-prediction. Thus optimal signal extraction with confounding dynamics generates fluctuations where the full-information and exogenously imposed noise counterparts generate none. We now show how to embed these dynamics in a rational expectations model.

3 Model, Information, and Equilibrium

We now model confounding dynamics in a generic rational expectations formulation that permits many interpretations (e.g., monetary model, asset pricing model, etc.). We do this via dispersed information, which introduces well-known difficulties. We lay out a solution strategy that takes advantage of aforementioned mathematical properties and compare that strategy to alternative methodologies.

3.1 Model We consider models that are populated by a continuum of agents indexed by $i \in [0, 1]$. Let $\mu(i)$ be the density of agent i characterized by the information set at time t, denoted by Ω_{it}. We are interested in the class of models in which the individual optimal choice can be represented by the dynamic expectational difference equation,

$$\phi \mathbb{E}\big[X_{it+1}\big|\Omega_{it}\big] = \psi(L)X_{it}, \tag{3.1}$$

where

$$X_{it} \equiv \begin{pmatrix} x_{it} & y_t & \theta_{it} \end{pmatrix}^\top \tag{3.2}$$

Here $\phi \equiv [\phi_x \ \phi_y \ \phi_\theta]$, is a vector of coefficients, and $\psi(L) \equiv [\psi_x(L) \ \psi_y(L) \ \psi_\theta(L)]$, is a vector of square-summable lag polynomials in non-negative powers of L. x_{it} is the choice variable under the control of the individual agent i; y_t is an endogenous aggregate variable that agents take as given, and θ_{it} is an exogenous stochastic process specified as the sum of an aggregate component θ_t and an i.i.d. individual component v_{it}. Formally

$$\theta_{it} = \theta_t + v_{it}, \quad \text{where} \quad \theta_t = A(L)\varepsilon_t, \tag{3.3}$$

with $v_{it} \sim \mathcal{N}(0, \sigma_v)$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$, and $A(L)$ is a square-summable polynomial in non-negative powers of L. To close the model we need to specify a relationship between the distribution of x_{it} across agents, and the aggregate y_t. We thus posit that

$$\gamma(L) \int_0^1 X_{it}\mu(i)di = 0, \tag{3.4}$$

where $\gamma(L) \equiv [\gamma_x(L) \ \gamma_y(L) \ \gamma_\theta(L)]$, is a vector of square-summable finite-degree lag polynomials in non-negative powers of L, and we assume $\gamma_x(L) \neq 0$.[8] As we proceed with the analysis it will be useful to think of equation (3.1) as representing a demand or supply schedule for agent i, and (3.4) as the relative market clearing condition. However, the specific form depends on the particular application at hand. As we show in Section 3.3, this setup nests incomplete information models typical of the macro and finance literatures.

The expectational difference equation (3.1) is a dispersed information version of the system originally considered by Blanchard and Kahn (1980), and subsequently studied by Uhlig (1999), Klein (2000) and Sims (2002), among others. Dispersed information implies that individual expectations are heterogeneous, which implies that the aggregation in (3.4) will result in taking an average of expectations. In particular, model (3.1)-(3.4) can accommodate both average expectations of aggregate variables and average expectations of individual variables.

3.2 Information In our dispersed information setup, we assume that the information set Ω_{it} of an arbitrary agent i at time t consists of the smallest closed subspace generated by the history of the random variable $\theta_i^t \equiv \{\theta_{it}, \theta_{it-1}, ...\}$, and the history of the aggregate variable $y^t = \{y_t, y_{t-1}, ...\}$. Specifically, $\Omega_t^i = \theta_i^t \vee y^t$, where the operator $\vee$ denotes the span (i.e., the smallest closed subspace

[8] We make this assumption in order to keep the connection between (3.1) and (3.4) non-trivial. Allowing for $\gamma_x(L) = 0$, would imply that y_t is directly determined by the process θ_t, and, as a consequence, it would enter (3.1) as an exogenous variable, essentially duplicating the role of θ_t in that equation.

which contains the subspaces) generated by the sequences θ_i^t and y^t. This notation simply suggests that expectations will be taken optimally; i.e., they will be consistent with the prediction formulas discussed in Section 2.2. In a multivariate moving-average setting, the invertible representation achieved via canonical factorization is the smallest closed subspace containing the observables, θ_i^t and y^t (see Hoffman (1962)).

Given (3.1), x_{it} will be a function of the history of idiosyncratic innovations, v_{it}, and the aggregate innovations, ε_t, namely

$$x_{it} = X(L)\varepsilon_t + V(L)v_{it}. \tag{3.5}$$

In addition, aggregation implies that y_t is only a function of aggregate innovations, so that

$$y_t = Y(L)\varepsilon_t. \tag{3.6}$$

The signal structure can be thus represented as

$$\begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix} = \Gamma(L) \begin{pmatrix} \sigma_\varepsilon^{-1}\varepsilon_t \\ \sigma_v^{-1}v_{it} \end{pmatrix}, \qquad \Gamma(L) = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ Y(L)\sigma_\varepsilon & 0 \end{bmatrix}. \tag{3.7}$$

We point out that our information set is in line with the typical information set assumed in the dispersed information rational expectations literature: we provide agents with both an *exogenous* signal about the aggregate unobserved state (θ_{it}), and an *endogenous* signal that is determined in equilibrium (y_t). The analytical convenience of the signal structure (3.7), for our purposes, is that the invertibility of the matrix $\Gamma(L)$ hinges only upon the zeros of $Y(L)$. At the same time, the structure imposes analytical discipline that is uncommon in the literature: the endogenous signal y_t can reveal perfectly the underlying state, under the appropriate parametrization of model (3.1)-(3.4). Thus, the theorems below must establish both the degree to which information remains incomplete in equilibrium, along with the more standard existence and uniqueness conditions.

3.3 Examples In this section we present four applications that can be cast into our model specification. The list is by no means exhaustive.

Example 1: Real Business Cycle with Capital. In a standard real business cycle model with capital, in presence of dispersed information about the aggregate productivity shock and incomplete insurance markets, the linear dynamics of capital around the steady state can be expressed as

$$\alpha\beta\mathbb{E}_{it}\big(k_{it+2}\big) + (1-\alpha\beta)\mathbb{E}_{it}\big(r_{t+1}\big) - \mathbb{E}_{it}\big(a_{it+1}\big) = \alpha(1+\beta)k_{it+1} - \alpha k_{it} - a_{it}, \tag{3.8}$$

which is the standard second-order difference equation in capital and where

$$r_{t+1} = \int_0^1 a_{it+1}\mu(i)di - (1-\alpha)\int_0^1 k_{it+1}\mu(i)di, \tag{3.9}$$

is the market-clearing rental rate for capital. Here β is the discount factor, α is the capital share in the Cobb-Douglas output good technology, and a_{it} is the individual productivity shock. Model (3.8)-(3.9) maps into (3.1)-(3.4) by setting $x_{it} = k_{it+1}$, $y_t = r_t$, $\theta_{it} = a_{it}$, and

$$\phi = \begin{pmatrix}\alpha\beta & (1-\alpha\beta) & -1\end{pmatrix}, \psi(L) = \begin{pmatrix}\alpha(1+\beta) - \alpha L & 0 & -1\end{pmatrix}, \gamma(L) = \begin{pmatrix}(1-\alpha)L & 1 & -1\end{pmatrix}.$$

Example 2: Calvo Pricing and New Keynesian Phillips Curve. Nimark (2008) considers a dispersed information version of the optimal Calvo pricing problem that micro-founds the New Keynesian Phillips Curve popularized Woodford (2003b) and Galì (2008). Nimark (2008) shows that the optimal price for an optimizing firm is

$$p^*_{it} = (1-\beta\vartheta)(p_t + mc_{it}) + \beta\vartheta\mathbb{E}_{it}\big(p^*_{it+1}\big) \tag{3.10}$$

where p_t is the aggregate price level, defined as $p_t = \vartheta p^*_t + (1-\vartheta)p_{t-1}$, with $p^*_t \equiv \int_0^1 p^*_{it}\mu(i)di$, and mc_{it} is the individual marginal cost at time t specified as $mc_{it} = mc_t + v_{it}$ so that $\int_0^1 mc_{it}\mu(i)di = mc_t$. The parameter β is the discount factor for price setters, while ϑ measures the probability of resetting ones' price in a given period. Define $p_{it} \equiv \vartheta p^*_{it} + (1-\vartheta)p_{t-1}$, which maintains $\int_0^1 p_{it}\mu(i)di = p_t$. The individual and aggregate price level dynamics can then be written as,

$$\beta\vartheta\mathbb{E}\big(p_{it+1}\big|\Omega_{it}\big) = p_{it} - \vartheta\big(1 - 2\vartheta\beta + \beta\big)p_t + (1-\vartheta)p_{t-1} - \vartheta(1-\beta\vartheta)mc_{it}. \tag{3.11}$$

with

$$p_t = \int_0^1 p_{it}\mu(i)di. \tag{3.12}$$

Equations (3.11)-(3.12) maps into (3.1)-(3.4) by setting $x_{it} = p_{it}$, $y_t = p_t$ $\theta_{it} = mc_{it}$, and

$$\phi = \begin{pmatrix}\beta\vartheta & 0 & 0\end{pmatrix}, \psi(L) = \begin{pmatrix}1 & -\vartheta\big(1-2\vartheta\beta+\beta\big) + (1-\vartheta)L & -\vartheta(1-\beta\vartheta)\end{pmatrix}, \gamma(L) = \begin{pmatrix}1 & -1 & 0\end{pmatrix}.$$

In the presence of dispersed information a compact representation of the New Keynesian Phillips Curve cannot be obtained. However, once a solution for p_t is derived from (3.11)-(3.12), inflation dynamics are immediately given by $\pi_t = p_t - p_{t-1}$.

Example 3: Dynamic Asset Pricing. Singleton (1987) presents a dynamic asset pricing model motivated by the market microstructure of the U.S. bond market, which features a competitive, Walrasian market structure with a single security that is traded among speculative investors and non-speculative or liquidity traders at the price p_t.[9] The security is assumed to pay a constant coupon every period, which we normalize to zero. Purchases of the security are financed by borrowing at the constant rate r, and the wealth of investor i evolves according to $w_{it+1} = z_{it}p_{t+1} - (1+r)(z_{it}p_t - w_{it})$. The ith investor is assumed to have a one-period investment horizon and to rank alternative investment strategies according to the utility $\mathbb{E}_{it}\big[-\exp(-\varrho w_{it+1})\big]$, where ϱ is the constant coefficient of

[9] Several papers have since used a similar setup to study a broad range of asset pricing issues (e.g., Bacchetta and van Wincoop (2006)).

absolute risk aversion. Singleton shows that the demand schedule for the risky asset takes the form

$$z_{it} = \frac{1}{\varrho\nu}\mathbb{E}_{it}(p_{t+1}) - \frac{1+r}{\varrho\nu}p_t, \tag{3.13}$$

where ν is the variance of p_{t+1} and is set to be an exogenous constant. Singleton (1987) assumes that the net supply of the asset, denoted by s_t, is given by

$$s_t = f_t + \vartheta p_t. \tag{3.14}$$

The shock to net asset supply f_t arises from nonspeculative traders (such as the U.S. Treasury, the Federal Reserve, financial intermediaries), that attempt to satisfy macroeconomic objectives for technical reasons related to the intermediation process. Nonspeculative traders are assumed to respond positively to an increase in prices; thus $\vartheta > 0$. Investors in setting their strategy z_{it} are assumed to receive a private signal $f_{it} = f_t + v_{it}$ about the shock to the net asset supply. Market clearing is therefore given by

$$\int_0^1 z_{it}\mu(i)di = s_t. \tag{3.15}$$

Model (3.13)-(3.15) maps into (3.1)-(3.4) by setting $x_{it} = z_{it}$, $y_t = p_t$, $\theta_{it} = f_{it}$, and

$$\phi = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \psi(L) = \begin{pmatrix} \varrho\nu & 1+r & 0 \end{pmatrix}, \gamma(L) = \begin{pmatrix} 1 & -\vartheta & -1 \end{pmatrix}.$$

Example 4: Classical Monetary Models of Inflation. In classical monetary models of inflation, money demand takes the form popularized by Cagan (1956),

$$m_{it} - p_t = -\alpha\big(\mathbb{E}_{it}(p_{t+1}) - p_t\big), \tag{3.16}$$

where m_{it} is nominal money demand by agent i, p_t is an aggregate price index, and $\alpha > 0$. The money supply M_t is assumed to possess persistent dynamics specified as

$$M_t = \rho M_{t-1} + s_t, \tag{3.17}$$

where $s_t = A(L)\varepsilon_t$ is a money supply shock process. The money market clearing condition is then

$$M_t = \int_0^1 m_{it}\mu(i)di \tag{3.18}$$

Agents are assumed to receive a private signal, $s_{it} = s_t + v_{it}$, about the money supply shock. Equations (3.16)-(3.18) map into (3.1)-(3.4) by setting $x_{it} = m_{it}$, $y_t = p_t$, $\theta_{it} = s_{it}$, and

$$\phi = \begin{pmatrix} 0 & -\alpha & 0 \end{pmatrix}, \psi(L) = \begin{pmatrix} 1 & -(1+\alpha) & 0 \end{pmatrix}, \gamma(L) = \begin{pmatrix} 1-\rho L & 0 & -1 \end{pmatrix}.$$

3.4 Equilibrium Definition Uncertainty is assumed to be driven by Gaussian innovations, which, together with linearity, implies that conditional expectations are computed as optimal linear projections. We thus have

$$\mathbb{E}\big(X_{it+1}\big|\Omega_{it}\big) = \mathcal{P}\big[X_{it+1}\big|\Omega_{it}\big], \tag{3.19}$$

and can apply the prediction formulas in Section 2.1 to compute conditional expectations. We are now ready to define a Rational Expectations Equilibrium for model (3.1)-(3.4).

Definition REE. *A Rational Expectations Equilibrium (REE) is a stochastic process for* $\{X_{it}, i \in [0,1]\}$ *and a stochastic process for the information sets* $\{\Omega_{it}, i \in [0,1]\}$ *such that: (i) each agent* i*, given her information set, forms expectations according to* (3.19)*; (ii)* $\{X_{it}, i \in [0,1]\}$ *satisfies conditions* (3.1)-(3.4)*.*

The REE can be summarized by two statements: (a) given a distribution of information sets, there exists a market clearing distribution $\{X_{it}, i \in [0,1]\}$ determined by each agent i's optimal prediction conditional on the information sets; (b) given a distribution $\{X_{it}, i \in [0,1]\}$, there exists a distribution of information sets that provides the basis for optimal prediction. Both statements (a) and (b) must be satisfied by the same distribution $\{X_{it}, i \in [0,1]\}$ and the same distribution of information sets *simultaneously* in order to satisfy the requirements of a REE. This dual fixed point condition is standard in rational expectations with potentially heterogeneously informed agents and when endogenous variables convey information [see, Radner (1979) as an early example].

3.5 Weighted Sum of Expectations Before discussing our solution methodology, we give a brief overview of an alternative approach to solve model (3.1)-(3.4), which consists of two steps. The first step is to iteratively substitute the endogenous variables x_{it+j} and y_{t+j} forward by leading (3.1) j periods forward and aggregating over agents. The end result is an expression for x_{it} and y_t, that are a function of expectations of the exogenous variable θ_t at all future horizons. The second step is then to compute those expectations, which is non-trivial due to the fact that the law of iterated expectations may not be operational. Nearly all competing approaches that take this approach rely on numerics to calculate these expectations [Nimark (2010), Melosi (2016)].

Specifically, consider the univariate case, where $y_t = \int_0^1 x_{it}\mu(i)di$, setting $\phi_\theta = 0$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, and $\psi_\theta(L) = -1$. Next, bring (3.1) one period forward to obtain an expression for x_{it+1} and aggregate to get the analogue expression for y_{t+1}. Substitute both expressions back into (3.1), which now will contain x_{it+2} and y_{t+2}. Iterating on this procedure and aggregating over agents, the expression for y_t becomes[10]

$$y_t = \sum_{j=1}^{\infty} \mathbb{P}^j_{\phi_x}\Big[(\phi_x + \phi_y)^j \bar{\mathbb{E}}^j(\theta_{t+j})\Big] \tag{3.20}$$

Here $\bar{\mathbb{E}}^j(\theta_{t+j})$ stands for the j^{th} order average expectation of θ_{t+j}, with the convention that $\bar{\mathbb{E}}^0(\theta_t) = \theta_t$. Disregarding for a moment the operator $\mathbb{P}^j_{\phi_x}$, equation (3.20) shows that y_t can be represented

[10] We provide the details of the derivation in Appendix B.

as a weighted sum of average expectations of higher order about the future realizations of the exogenous process θ_t. The higher the values of ϕ_x and ϕ_y, the higher the relative weight of higher order thinking. The operator $\mathbb{P}^j_{\phi_x}$ operates on the order of expectations j by reducing some of the higher order compounding depending on the position of ϕ_x in the terms of the polynomial $(\phi_x + \phi_y)^j$. To see why the order might need to be reduced, consider the expression for $\phi_x \mathbb{E}_{it}(x_{it+1})$ which contains the term $\phi_x \mathbb{E}_{it+1}(x_{it+2})$, whose expression in turn contains θ_{t+2}. It follows that the law of iterated expectations (LIE) applies in this context so that $\phi_x^2 \mathbb{E}_{it}\mathbb{E}_{it+1}(\theta_{t+2}) = \phi_x^2 \mathbb{E}_{it}(\theta_{t+2})$, and aggregation implies $\phi_x^2 \bar{\mathbb{E}}_t(\theta_{t+2})$ for $j = 2$ in (3.20). Intuitively, in each round of the iterative substitutions to achieve representation (3.20) there are terms where agent i is taking expectations of both her own future expectations and of future average expectations. The law of iterated expectations applies to the former, so that the order of expectations is reduced, but not to the latter.[11] It should be evident at this point that the second step required by the canonical approach—computing closed form solutions for the expectations of arbitrary order—is a daunting task under dispersed information. As already remarked and discussed thoroughly in the next section, we approach the solution to (3.20) from a completely different angle.

3.6 Solution Methodology Our aim is to characterize an equilibrium for model (3.1)-(3.4) with confounding dynamics. The critical requirement for confounding dynamics to emerge is that the information matrix $\Gamma(L)$ must be non-invertible at a $\lambda \in (-1, 1)$; i.e., incomplete information must be present in equilibrium. However, there is no guarantee that this condition will hold. Following Townsend (1983a), our approach is to formulate a guess for the endogenous variables that follows a generic polynomial in the underlying shocks, and then derive conditions on the *exogenous* parameters that yield non-invertibility ($\lambda \in (-1, 1)$) in equilibrium. The following steps describe our procedure.

1. Specify the guesses for x_{it} and y_t as generic polynomials of underlying shocks

$$x_{it} = X(L)\varepsilon_t + V(L)v_{it}, \quad \text{and} \quad y_t = Y(L)\varepsilon_t. \tag{3.21}$$

   with

$$Y(\lambda) = 0, \quad \text{for} \quad \lambda \in (-1, 1). \tag{3.22}$$

2. Given the signal matrix $\Gamma(L)$, obtain the canonical factorization form $\Gamma^*(L)$ under (3.22).

3. Use the canonical factorization $\Gamma^*(L)$ in formula (2.4) together with the guesses (3.21)-(3.22) to obtain the conditional expectations in (3.1).

4. Aggregate over agents according to (3.4) and use the relationship between $X(L)$ and $Y(L)$ to substitute $X(L)$ with $Y(L)$ in (3.1). Both the right hand side and the left hand side will now

[11] Mechanically, whether LIE applies or not at each iteration depends on the position of ϕ_x in the coefficients of the polynomial $(\phi_x + \phi_y)^j$, i.e. on the set of permutations of size j of ϕ_y and ϕ_x with repetition. For instance, for the case of $j = 2$, the set of terms that multiply ψ_θ in (3.20) are $(\phi_y^2 + \phi_y\phi_x)\bar{\mathbb{E}}_t\bar{\mathbb{E}}_{t+1}(\theta_{t+2}) + (\phi_x\phi_y + \phi_x^2)\bar{\mathbb{E}}_t(\theta_{t+2})$. For more details on this please see Appendix B.

be lag polynomial operators in ε_t and v_{it}, and will thus provide the fixed point conditions for $Y(L)$ and $V(L)$.

5. Derive conditions on exogenous parameters so as to ensure that the solution exists and is unique, and that there exists a $|\lambda| < 1$, verifying (3.22). Once $Y(L)$ is solved for, use (3.4) to recover $X(L)$.

Note that at no point in the solution procedure one needs to worry about higher-order expectations. The so-called "higher-order thinking" that complicates the iterative approach is implicit in how the guess (3.21) combines with the information matrix $\Gamma(L)$ to provide a closed form for the first order expectations in (3.1). By guessing a generic lag polynomial, the higher-order beliefs are built into the guess and we do not have to track these terms explicitly. (Although higher-order beliefs can be backed out of the solution in closed form.) This is why Townsend (1983a) forcefully advocated for the solution methodology outlined above.

4 Equilibrium with Confounding Dynamics

This section establishes the main result of the paper: the existence of a rational expectations equilibrium with confounding dynamics in a dispersed information environment.

4.1 Full Information Benchmark We consider first the case of Full Information to establish a useful benchmark and to show, in the simplest of settings, how our solution methodology works. We define Full Information as the case when every agent is endowed with perfect knowledge of the aggregate and her own idiosyncratic innovations history up to time t. Denoting the full information set by $\tilde{\Omega}_{it}$, the set is formally specified as

$$\tilde{\Omega}_{it} = v_i^t \vee \varepsilon^t \tag{4.1}$$

Here, and in the following analysis, we assume that agents always observe the history of the endogenous aggregate y_t up to time t, and that they know that the equilibrium relationship is given by (3.1)-(3.4). When information is full, all agents share the same expectations about the future value of the exogenous process θ_t. As a consequence, the entire structure of higher-order expectations in (3.20) collapses to the common first-order expectation so that

$$y_t = \sum_{j=0}^{\infty} (\phi_x + \phi_y)^j \mathbb{E}_t(\theta_{t+j}) \tag{4.2}$$

Hansen and Sargent (1980) worked out a formula to express the discounted sum of future expectations—such as the one in (4.2)—in closed form, which since then has been known as the Hansen-Sargent formula. Here we show how to derive the Hansen-Sargent formula by applying the methodology of Whiteman (1983).[12] Our main theorem extends the Hansen-Sargent formula to models with

[12] Whiteman (1983) provides a rigorous treatment of solving linear rational expectation models using the space of analytic functions. We rely on his theorems to establish existence and uniqueness of equilibria and provide an

incomplete information.

Throughout the full information analysis we will maintain that $\phi_\theta = 0$, $\psi_\theta(L) = -1$, and consider market clearing (3.4) with $\gamma_\theta(L) = 0$, so that

$$\tilde{\gamma}_y(L) y_t = \int_0^1 x_{it} \mu(i) di \tag{4.3}$$

Here $\tilde{\gamma}_y(L) \equiv \frac{\gamma_y(L)}{\gamma_x(L)}$, and we assume that $\gamma_x(0) \neq 0$, so that $\tilde{\gamma}_y(0)$ is well defined.[13]

We begin by guessing that the solution takes the form, $x_{it} = \mathcal{X}(L)\varepsilon_t + \mathcal{V}(L)v_{it}$, and $y_t = \mathcal{Y}(L)\varepsilon_t$, where $\mathcal{X}(L)$, $\mathcal{V}(L)$ and $\mathcal{Y}(L)$ are square-summable lag polynomial in non-negative powers of L. Under full information, direct application of the Wiener-Kolmogorov formula (2.5) provides expressions for the relevant expectational terms,

$$\mathbb{E}_{it}(x_{it+1}) = [\mathcal{X}(L) - \mathcal{X}(0)]L^{-1}\varepsilon_t + [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it}, \tag{4.4}$$

$$\mathbb{E}_{it}(y_{t+1}) = [\mathcal{Y}(L) - \mathcal{Y}(0)]L^{-1}\varepsilon_t, \tag{4.5}$$

$$\mathbb{E}_{it}(\theta_{t+1}) = [A(L) - A(0)]L^{-1}\varepsilon_t. \tag{4.6}$$

Substituting these expressions into equation (3.1), invoking $\mathcal{X}(L) = \tilde{\gamma}_y(L)\mathcal{Y}(L)$, and aggregating over agents, we obtain an equation featuring $\mathcal{Y}(L)$, while $\mathcal{V}(L)$ is washed out by the aggregation process. After substitution, the ε_t terms can be dropped, both sides can be multiplied by L, and the expression can be rearranged to solve for $\mathcal{Y}(L)$ so that

$$\mathcal{Y}(L) = -\frac{(\phi_x \tilde{\gamma}_y(0) + \phi_y)\mathcal{Y}(0) + A(L)L}{\Phi(L)}, \tag{4.7}$$

where,

$$\Phi(L) \equiv \phi_x \tilde{\gamma}_y(L) + \phi_y - (\psi_x(L)\tilde{\gamma}_y(L) + \psi_y(L))L. \tag{4.8}$$

$\Phi(L)$ is the characteristic polynomial that drives the auto-regressive behavior and the stationarity of y_t. Expression (4.7) is not a solution because it features the endogenous constant $\mathcal{Y}(0)$ on both sides, and, if we evaluate both sides at $L = 0$, the term $\mathcal{Y}(0)$ drops out and cannot be solved for. As shown in Whiteman (1983), this indeterminacy is pinned down by the requirement that $\mathcal{Y}(L)$ be square-summable (i.e. stationary), which is equivalent to the denominator polynomial of $\Phi(L)$ not having any roots inside the unit circle. Thus, the constant $\mathcal{Y}(0)$ must be set to remove roots inside the unit circle in $\Phi(L)$.

Appendix A.2 shows that there are three cases to consider: [i.] if there are no roots of $\Phi(L)$ inside the unit circle, then an infinite number of equlibria exist because $\mathcal{Y}(0)$ cannot be set uniquely. [ii.] if there are multiple roots of $\Phi(L)$ inside the unit circle, then no stationary equilibria exists. [iii.] and if there is exactly one root of $\Phi(L)$ inside the unit circle, then $\mathcal{Y}(0)$ can be set uniquely

overview of his approach in the appendix.

[13] The solution under the general specification (3.1)-(3.4) is not instructive for our purposes here and has been relegated to Appendix A.2.

and a unique equilibrium exists. Thus we make the following assumption.

Assumption ($\mathcal{S}$). *The polynomial $\Phi(L)$ has exactly one root inside the unit circle.*

It is important to note that Assumption ($\mathcal{S}$) is not a special case. It is the standard assumption necessary to yield a unique rational expectations equilibrium (e.g., Sims (2002)) and it immediately implies that $\Phi(L)$ can be factorized as

$$\Phi(L) = (\zeta - L)\tilde{\Phi}(L), \tag{4.9}$$

where $|\zeta| < 1$, and $\tilde{\Phi}(L)$ has no roots inside the unit circle. Under assumption ($\mathcal{S}$), the constant $\mathcal{Y}(0)$ can be chosen so to introduce a root in the numerator polynomial of (4.7) that cancels the non-stationary root ζ at the denominator. To wit,

$$(\phi_x \tilde{\gamma}_y(0) + \phi_y)\mathcal{Y}(0) + A(\zeta)\zeta = 0. \tag{4.10}$$

Solving for $\mathcal{Y}(0)$ and substituting the expression into (4.7), one finally obtains the solution

$$\mathcal{Y}(L) = \frac{A(\zeta)\zeta - A(L)L}{\Phi(L)}, \tag{4.11}$$

where, by construction, the root ζ in the denominator is now canceled with a zero at ζ in the numerator.[14]

Equation (4.11) is an instance of the Hansen-Sargent formula. To better understand the formula, again let $\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, so that $\zeta = \phi_x + \phi_y$, and note that $1/(\zeta - L) = -L^{-1}(1 + \zeta L^{-1} + \zeta^2 L^{-2} + ...)$. After some manipulation, one can write $y_t = \mathcal{Y}(L)\varepsilon_t$ as

$$y_t = \sum_{j=0}^{\infty} \zeta^j \theta_{t+j} - A(\zeta) \sum_{j=1}^{\infty} \zeta^j \varepsilon_{t+j}. \tag{4.12}$$

Comparing this expression to (4.2) shows that the Hansen-Sargent formula turns the infinite sum of expectations about future θ_t's into the difference between the infinite sum of future θ_t's under perfect foresight (the first summation term in square brackets), minus the innovations to those future realizations that are not known at time t given the specified information set (the second summation term in square brackets). In this sense, it is a true prediction formula. It takes the best guess if all information were available to the agents and subtracts off the precise linear combination of unknown elements that minimizes the agent's forecast error.

We conclude this section by pointing out that, even though one does not need to solve for $\mathcal{V}(L)$ to figure out the solution for y_t (because all agents are equally informed), one can apply the same steps as above to obtain a closed form for $\mathcal{V}(L)$. The agent-specific component $\mathcal{V}(L)v_{it}$ determines the cross section distribution of x_{it}. The characteristic polynomial that drives the autoregressive

[14] In this sense, (4.11) can be simplified further to cancel the two roots. However, unless one specifies a specific process for $A(L)$, and an explicit form for the polynomials in $\Phi(L)$, such cancellation cannot be undertaken algebraically.

behavior of $\mathcal{V}(L)$ is $\phi_x(L) \equiv \phi_x - \psi_x(L)L$. In order to have the cross-sectional distribution well defined at any point in time, except possibly for the unit root limit, we assume the following.

Assumption (s). *The polynomial $\phi_x(L)$ has exactly one root inside the unit circle.*

We report the closed form solution to $\mathcal{V}(L)$ in Appendix A.2.

4.2 Equilibrium with Confounding Dynamics: Main Theorem In this section we state our main Theorem, which provides conditions under which a REE with Confounding Dynamics exists. As stated in Section 3.2, we specify the information set as

$$\Omega_{it} = x_i^t \vee y^t \tag{4.13}$$

Agents thus observe the entire history of the exogenous process θ_{it} up to time t, together with the history of the aggregate variable y_t. In addition, the signal structure (2.2) and the model equations (3.1)-(3.4) are both common knowledge across agents.

Given our definition of confounding dynamics 2.1 and signal structure (2.2), we must find restrictions on exogenous processes that ensure for $y_t = Y(L)\varepsilon_t$, there exists a $\lambda \in (-1, 1)$ such that

$$Y(\lambda) = 0 \tag{4.14}$$

The following theorem derives conditions for such a λ to exist.

Theorem 1. *Consider model (3.1)-(3.4) with Assumptions ($\mathcal{S}$) and (s). Let the information sets be specified as in (4.13). There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form, $y_t = Y(L)\varepsilon_t$, with*

$$Y(L) = \mathcal{Y}(L) - (1-\tau)(1-\lambda^2)\frac{\mathcal{A}(\lambda)}{(1-\lambda L)\tilde{\Phi}(L)}, \tag{4.15}$$

if, and only if, there exists a $\lambda \in (-1, 1)$ that solves

$$\mathcal{Y}(\lambda)\tilde{\Phi}(\lambda) = (1-\tau)\mathcal{A}(\lambda), \tag{4.16}$$

where $\mathcal{Y}(L)$ is the full information solution, $\tau \equiv \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2+\sigma_v^2}$, and $\mathcal{A}(\lambda)$ is a function of λ that depends only on exogenous parameters.

Proof. See Appendix A.3. □

Theorem 1 provides necessary and sufficient conditions for an equilibrium with confounding dynamics to exist. The existence condition is given by (4.16) while the functional form of the equilibrium is (4.15). The intuition underlying the existence condition follows directly from the signal extraction analysis of Section 2.2, where we show that if there is a univariate moving average process, $s_t = (\lambda + L)\varepsilon_t$, the root of the polynomial in L must lie inside the unit circle $|\lambda| < 1$ for confounding dynamics to hold. Theorem 1 applies the same concept as noted by (4.14), only now to an endogenous variable $Y(L)\varepsilon_t$. Therefore, we use the equilibrium restrictions on $Y(\cdot)$ to connect it to a purely exogenous process, which yields the RHS of (4.16). That is, in proving Theorem 1, we show that the functional form of the equilibrium polynomial $Y(L)$ is given by (4.15). Therefore, evaluating the right-hand side of (4.15) at λ and setting it equal to zero, is the necessary restriction for confounding dynamics. This is precisely what (4.16) achieves. Sufficiency of confounding dynamics requires that there exists a solution to this equation and that solution must lie inside the unit circle. If this restriction is not satisfied, the solution is given by the full-information equilibrium and the standard Hansen-Sargent formula (4.12). Below we examine Theorem 1 within the context of a Real Business Cycle model and provide further intuition regarding restriction (4.16).

The form of (4.15) is intuitive when contrasted with the full information counterpart. As noted above, the standard Hansen-Sargent formula (4.12) subtracts off the particular linear combination of *future* values of ε_t that minimize the agent's forecast error. As described in Section 2.2, confounding dynamics implies that a particular linear combination of *past* values of ε_t are never revealed to the agent. In order to make a direct comparison to the Hansen-Sargent formula (4.12), set $\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, $\phi_\theta = 0$ and $\psi_\theta(L) = -1$. According to Theorem 1, the solution under confounding dynamics can be written as

$$y_t = \sum_{j=0}^{\infty} \zeta^j \theta_{t+j} - A(\zeta) \sum_{j=1}^{\infty} \zeta^j \varepsilon_{t+j} - (1-\tau)(1-\lambda^2)\mathcal{A}(\lambda) \sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}. \tag{4.17}$$

The first two components on the right-hand side are exactly the Hansen-Sargent formula. The third component— represented by the weighted sum $\sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}$—arises due to confounding dynamics and is similar to the prediction formula of Section 2.2. Agents do not observe the linear combination of shocks weighted by λ. Conditioning down implies that this linear combination will (optimally) be subtracted from the Hansen-Sargent full-information equilibrium. The extent to which the unknown past matters depends on (i) the imprecision of the private signal θ_{it}, measured by $1-\tau$; (ii) the imprecision stemming from confounding dynamics, measured by $1-\lambda^2$; (iii) the fixed point constant $\mathcal{A}(\lambda)$. As in Section 2.2, λ completely parameterizes the size of the forecast error.

4.3 Exogenous Noise As discussed in Section 2.1, there are two ways to preserve heterogeneous information in equilibrium—by continually adding exogenous noise until the noise terms overwhelm all signals and/or by proving that there exists a zero inside the unit circle of the equilibrium as is done in Theorem 1. These categories are not mutually exclusive. Combinations of the two can certainly exist. However, we now show that the standard way of introducing exogenous noise will *not* lead to the characteristic over- and under-reaction of the impulse response associated with confounding dynamics.[15] We then demonstrate how a small change in the exogenous process can deliver these dynamics. For transparency, we work within the stylized version of the generic rational

[15]Superimposing exogenous noise is a common practice in most of the recent (and past) literature on dispersed information and aggregate fluctuations [e.g., Grossman and Stiglitz (1980), Wang (1993), Makarov and Rytchkov (2012), Angeletos and La'O (2013)].

expectations model [$\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, so that $\zeta = \phi_x + \phi_y$].

Assume that the information set of an arbitrary agent i still contains the private signal $\theta_{it} = \theta_t + v_{it}$, where $\theta_t = A(L)\varepsilon_t$ but now let the endogenous variable be observed with noise

$$\tilde{y}_t = y_t + \tilde{\eta}_t. \tag{4.18}$$

The noise $\tilde{\eta}_t$ is assumed to be of the form $\tilde{\eta}_t = U(L)\eta_t$, where $U(L)$ is a ratio of two lag polynomials in non-negative powers of L with zeros outside the unit circle, and η_t is i.i.d. Gaussian with distribution $N(0, \sigma_\eta)$. Following our solution strategy, we posit a candidate solution $y_t = Q_\varepsilon(L)\varepsilon_t + Q_\eta(L)\eta_t$. The following corollary characterizes analytically a rational expectations equilibrium for the exogenous noise economy.

Theorem 2. *Consider model (3.1)-(3.4) with Assumptions ($\mathcal{S}$) and (s) and let $\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, so that $\zeta = \phi_x + \phi_y$. Let the information sets be specified as $\Omega_{it} = x_i^t \vee \tilde{y}^t$. Define the relative signal-to-noise ratios $\tau_\eta \equiv \frac{\sigma_v^2\sigma_\varepsilon^2}{\sigma_\eta^2\sigma_v^2+\sigma_v^2\sigma_\varepsilon^2+\sigma_\eta^2\sigma_\varepsilon^2}$ and $\tau_v \equiv \frac{\sigma_\eta^2\sigma_\varepsilon^2}{\sigma_\eta^2\sigma_v^2+\sigma_v^2\sigma_\varepsilon^2+\sigma_\eta^2\sigma_\varepsilon^2}$, and let $U(L) = \lambda(L)$ where $\lambda(L) \equiv (LA(L) - \zeta\tau_v A(\zeta\tau_v))/(L - \zeta\tau_v)$.*

If the processes

$$\lambda(z) \ \text{ and } \ \lambda(z) + \zeta\tau_\eta \frac{\lambda(z) - \lambda(\zeta)}{z - \zeta} \tag{4.19}$$

are invertible for all z inside the unit circle, then the rational expectations equilibrium does not yield confounding dynamics and is given by

$$y_t = \left(\lambda(L) + \zeta\tau_\eta \frac{\lambda(L) - \lambda(\zeta)}{L - \zeta}\right)\varepsilon_t + \zeta\tau_\eta \left(\frac{\lambda(L) - \lambda(\zeta)}{L - \zeta}\right)\eta_t \tag{4.20}$$

Proof. See Appendix A. □

The building block of Theorem 2 is the polynomial $\lambda(L)$, which takes the form of a Hansen-Sargent formula involving $A(L)$ and $\tau_v\zeta$. To understand its role, suppose that the public information $\tilde{y}_t$ is made uninformative so that $\tau_\eta \to 0$ (i.e. $\sigma_\eta \to \infty$). The equilibrium would then just be equal to $y_t = \lambda(L)\varepsilon_t$, which is the first term in (4.20) with τ_v equal to τ from Theorem 1. As soon as public information is made informative two additional terms appear, one which captures the additional information about ε_t transmitted by the public information, and the other that injects the public noise η_t into the equilibrium price. Note that the two terms enter the equilibrium price with the same dynamics, which is a consequence of the assumption $U(L) = \lambda(L)$.[16] This process is also characterized by a Hansen-Sargent formula involving $\lambda(L)$ and $\tau_\eta\zeta$ this time. When public information is made arbitrarily precise, i.e. $\sigma_\eta \to 0$ so that $\tau_\eta = 1$ then (4.20) corresponds to the full information equilibrium (4.11).

A comparison of the theorems reveals that the additional noise of Theorem 2 coming from (4.18)

[16]This assumption is due to Taub (1989) and it greatly simplifies the canonical factorization step (step 2) of the solution procedure.

implies that condition (4.16) of Theorem 1, which guarantees heterogeneous beliefs are preserved in equilibrium, is no longer necessary. In fact, (4.19) is an explicit assumption that there are no zeros inside the unit circle, which is the standard assumption in models with exogenous noise. In turn, this implies that the equilibrium cannot support confounding dynamics.

To see this more clearly, suppose that $A(L) = 1+\theta L$ with $|\theta| < 1$ to satisfy (4.19). Substituting into the equilibrium (4.20) yields

$$y_t = (1 + \theta\zeta(\tau_v + \tau_\eta))\varepsilon_t + \theta\varepsilon_{t-1} + \zeta\tau_\eta(1 + \theta)\eta_t. \tag{4.21}$$

The impulse response dynamics of y_t to a shock ε_t are entirely consistent with the optimal prediction formula associated with the standard signal extraction problem ($x_t = \varepsilon_t + \eta_t$) described in Section 2.2. The impulse is smaller than the full information counterpart but otherwise unchanged (i.e. it matches the dynamics of Figure 1). Adding confounding dynamics to this setup is straightforward as one would only need to relax assumption (4.19). If there were a zero inside the unit circle (e.g., $|\theta| > 1$), the equilibrium would again display the waves of optimism and pessimism of confounding dynamics.

5 Application: Business Cycle with Confounding Dynamics

In this section we apply our results to a model of business cycle fluctuations driven by productivity shocks. The economy consists of a continuum of islands indexed by $i \in [0, 1]$. Each island is inhabited by an infinitely-lived representative household, and by a representative firm, also indexed by i. Household i supplies labor services exclusively to firm i in a decentralized competitive labor market or, equivalently, workers cannot move across islands. Households supply labor inelastically to firms, and the labor supply is normalized to 1. Households own capital in the economy and rent it out to firms in a centralized spot market. Firms use capital and labor to produce output, which is supplied in a centralized competitive spot market. Households derive utility from consuming the output good. Output is produced by firm i according to a Cobb-Douglas technology with capital and labor inputs – with income shares α, and $1 - \alpha$ respectively, and total factor of productivity that is firm-specific and denoted by $e^{a_{it}}$, where

$$a_{it} = a_t + v_{it}.$$

The term a_t is common across all the islands, while v_{it} is a productivity component that is specific to island i. In what follows, we consider a log-linearized version of the model with full capital depreciation and logarithmic utility for tractability. The derivation of the log-linearization and the analysis of a more general version of the model is reported in the Online Technical Appendix. Household i sets consumption intertemporally according to the Euler equation

$$\mathbb{E}_{it}\big(c_{it} - c_{it+1} + r_{t+1}\big) = 0, \tag{5.1}$$

The intertemporal budget constraint is

$$(1-\beta\alpha)c_{it}+\alpha\beta k_{it+1}=(1-\alpha)w_{it}+\alpha r_t-\alpha k_{it} \tag{5.2}$$

where k_{it+1} is the capital stock that household i is carrying into period $t+1$, w_{it} is the wage rate, r_t is the rental rate of capital, and $\beta\in(0,1)$ is the subjective discount factor. On the firm's side, output is produced according to

$$y_{it}=a_{it}+\alpha h_{it}, \tag{5.3}$$

where h_{it} denotes the capital rented by firms producing in island i in period t. The demand for capital is

$$h_{it}=w_{it}-r_t, \tag{5.4}$$

while the zero-profits condition for firms imply that the wage rate is $w_{it}=y_{it}$. Using this last equality and (5.4) one can solve for w_{it} through (5.3)

$$w_{it}=\tfrac{1}{1-\alpha}(a_{it}-\alpha r_t), \tag{5.5}$$

which can be substituted into the budget constraint for the household. Market clearing for capital is specified as

$$\int_0^1 h_{it+1}\mu(i)di=\int_0^1 k_{it+1}\mu(i)di\equiv k_{t+1}. \tag{5.6}$$

Combining (5.3) and (5.4) and using (5.6), the rental rate for capital can be written as

$$r_t=a_t-(1-\alpha)k_t. \tag{5.7}$$

Using the household's budget constraint at t and at $t+1$ to get expressions for c_{it} and c_{it+1}, and leading (5.7) one period forward, one can finally substitute into the Euler equation (5.1) and obtain a second-order difference equation for capital k_{it+1} of the form

$$\alpha\beta\mathbb{E}_{it}\big(k_{it+2}\big)+(1-\alpha\beta)\mathbb{E}_{it}\big(r_{t+1}\big)-\mathbb{E}_{it}\big(a_{it+1}\big)=\alpha(1+\beta)k_{it+1}-\alpha k_{it}-a_{it}, \tag{5.8}$$

As remarked in Section 3.3, this equation maps into our general model by setting $x_{it}=k_{it+1}$ and $y_t=r_t$.

Full Information We begin by deriving the full information solution for aggregate capital. We specify the guess

$$k_{it+1}=\mathcal{K}(L)\varepsilon_t+\mathcal{V}(L)v_{it},$$

From (5.7), the rental rate of capital $r_t=\mathcal{R}(L)\varepsilon_t$ is immediately given by

$$\mathcal{R}(L)=A(L)-(1-\alpha)\mathcal{K}(L)L.$$

The information sets under full information correspond to, $\Omega_{it}=v_i^t\vee\varepsilon^t$. Following the steps outlined in Section 4.1,[17] the unique full information solution can be obtained as

$$\mathcal{K}(L) = \frac{\big(L-\alpha\beta\big)A(L) - \big(\zeta-\alpha\beta\big)A(\zeta)}{\Phi_K(L)}, \tag{5.9}$$

where

$$\Phi_K(L) = \alpha\beta - (1+\alpha^2\beta)L + \alpha L^2, \tag{5.10}$$

Given that α (capital's share of production) and β (subjective discount factor) are both less than one, (5.10) contains one root inside the unit circle and one outside. That is, Assumption ($\mathcal{S}$) holds and we may write $\Phi_K(L)$ as

$$\Phi_K(L) = \alpha(\zeta - L)(\tilde{\zeta} - L), \quad \text{with} \quad |\zeta| < 1 \quad \text{and} \quad |\tilde{\zeta}| > 1. \tag{5.11}$$

Confounding Dynamics We now characterize an equilibrium with confounding dynamics and begin with specifying information sets. Because households participate in two competitive markets every period – the labor market and the rental market for capital – they always observe the island-specific wage rate w_{it}, and the rental rate r_t. The observation of w_{it} and r_t implies that household i can always back out a_{it} at time t through equation (5.5). As a consequence, observing the prices of labor and capital is equivalent to the information set

$$\Omega_{it} = a_i^t \vee r^t \tag{5.12}$$

We also assume that households cannot observe the aggregate capital k_t, so to avoid the full revelation of a_t, and thus v_{it}, which would be implied by (5.7).[18] Following Theorem 1, existence of confounding dynamics requires the process for $r_t = R(L)\varepsilon_t$ has the following property,

$$R(\lambda) = 0, \tag{5.13}$$

for a $\lambda \in (-1, 1)$. A direct application of Theorem 1 leads to the following corollary.

Corollary 1. *Consider the Real Business Cycle model (5.6)-(5.8) and (5.11). Let the information sets be specified as in (5.12). There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form, $k_t = K(L)\varepsilon_t$, with*

$$K(L) = \mathcal{K}(L) - (1-\tau)(1-\lambda^2)\frac{\mathcal{A}_k(\lambda)}{(1-\lambda L)(\tilde{\zeta} - L)}, \tag{5.14}$$

[17] In Section 4.1 we solved for $\mathcal{Y}(L)$, which would correspond to $\mathcal{R}(L)$ in the present application, and left the solution of $\mathcal{X}(L)$ – $\mathcal{K}(L)$ here – as a corollary. In the present application that route is precluded by the fact that here $\gamma_x(L) = (1-\alpha)L$, which means $\gamma_x(0) = 0$. We thus take the alternative route, as showed in Appendix A.2: we solve for $\mathcal{K}(L)$ and leave the solution to $\mathcal{R}(L)$ as a straightforward extension.

[18] There are many other information structures that would preserve confounding dynamics in this setting and would be consistent with the general specification of Section 3.1.

Table 1: Existence of Equilibrium with Confounding Dynamics

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Noise-Signal (σ_v/σ_e)	1	1.5	1.5	1.5	2.5	3
Autocorrelation (ρ)	0.5	0.5	0.9	0.99	0.99	0.99
Confound. Dyn. (λ)	-1.39	-1.46	-0.95	-0.83	-0.78	Mult.

Existence of Equilibrium with Confounding Dynamics for Numerical Values of the noise-to-signal ratio σ_v/σ_e, and autocorrelation coefficient ρ. The rest of the parameters are set at $\beta = 0.985$, $\alpha = .33$. The entry "mult." means that there are multiple values of λ that are inside the unit circle.

if, and only if, there exists exactly one λ*, such that* $|\lambda| \in (0,1)$*, that solves*

$$\mathcal{R}(\lambda)(\lambda - \tilde{\zeta}) = (1-\alpha)(1-\tau)\mathcal{A}_k(\lambda)\lambda. \tag{5.15}$$

where $\tau \equiv \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2+\sigma_v^2}$*, and* $\mathcal{A}_k(\lambda)$ *is a function of* λ *that depends only on exogenous parameters.*

The functional forms of equations (5.14)–(5.15) have the same general structure as Theorem 1 (and same interpretation). In order to dig deeper into these conditions, we consider specific numerical examples. We assume that the process for a_t takes the $AR(1)$ form

$$a_t = \rho a_{t-1} + \varepsilon_t, \tag{5.16}$$

with $\rho \in [0,1]$, so that $A(L) = \frac{1}{1-\rho L}$.

Table 1 reports the endogenous values of λ, solved using (5.15), for several combinations of noise-to-signal ratios (σ_v/σ_e) and autocorrelation coefficients (ρ). Higher values of σ_v/σ_e correspond to a less informative private signal θ_{it}. Cases 1 and 2 show that an equilibrium with confounding dynamics does not exist for low values of both the noise-to-signal ratio (σ_v/σ_e) and autocorrelation coefficient (ρ). The equilibrium is fully revealing under this parameterization, and is given by the full-information equilibrium (5.9). As those values are increased, a solution $|\lambda| \in (0,1)$ to (5.15) appears. Recall that the lower is the absolute value of λ, the stronger is the incomplete information due to confounding dynamics. Cases 3, 4 and 5 show that confounding dynamics are stronger when the precision of private information is lower and persistence of the unknown aggregate shock is higher. Finally, Case 6 shows that, as private information is decreased further, multiple solutions are possible to (5.15), which means that the fixed point of Proposition 1 is no longer valid and a new guess with multiple λ's inside the unit circle is in order. While not discussed in the analysis, multiple λ's inside the unit circle have the opportunity to generate even more interesting dynamics relative to the single λ case.

Using (5.16) in (5.9), the full information equilibrium takes the $AR(2)$ form given by

$$\mathcal{K}(L) = \frac{1}{\alpha(1-\rho\zeta)} \frac{1}{(1-\rho L)(\tilde{\zeta} - L)}. \tag{5.17}$$

When an equilibrium with confounding dynamics exists, the solution takes the $ARMA(3,1)$ form

$$K(L) = \mathcal{K}(L) \left(\frac{1+\kappa - (\lambda + \kappa\rho)L}{1-\lambda L} \right). \tag{5.18}$$

where, $\kappa \equiv -\alpha(1-\tau)(1-\lambda^2)\mathcal{A}_k(\lambda)(1-\rho\zeta)$. Figure 2 shows the response of capital k_{t+1}, aggregate output y_t, aggregate consumption c_t, and the interest rate r_t, to a persistent unitary positive shock to aggregate productivity a_t. In interpreting the shape of the responses, it is useful to recall the following aggregate relationships,

$$y_t = a_t + \alpha k_t, \tag{5.19}$$

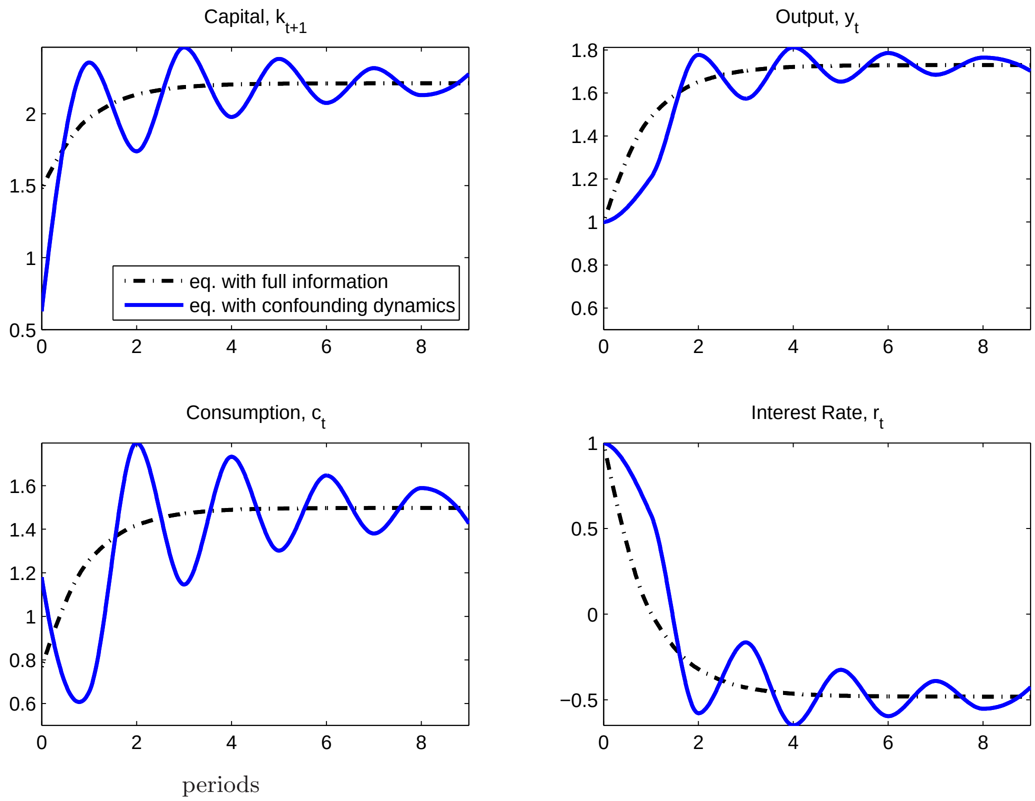
$$c_t = \frac{1}{1-\alpha\beta} y_t - \frac{\alpha\beta}{1-\alpha\beta} \alpha k_{t+1}, \tag{5.20}$$

together with (5.7).

In the full information equilibrium, represented by the dashed-lines, all variables display the standard real business cycle behavior. Capital increases at impact because a higher marginal return is anticipated for the future, and it climbs smoothly towards a new persistent higher level. Output jumps at impact because capital is fixed, while productivity has increased. Consumption also increases because, under the model's parametrization, it is optimal not not to use all the higher output for investment. Finally, the interest rate at impact jumps because supply is fixed, while demand from firms is higher. After the impact period, supply of capital is increased due to investment, which puts a downward pressure on the interest rate for all the subsequent periods, as capital, output and consumption all converge slowly to the new persistent higher level.

Under confounding dynamics, represented by the blue solid line, the overall long term pattern of the response is anchored around the full information benchmark, but the short term dynamics are markedly different. At impact, agents are not sure whether the increase in productivity they observe through wages is due to an aggregate productivity increase, or just an increase that is specific to their home island. As a consequence, as they try to parse out the effect, their investment demand is toned down, compared to the full information case – an increase of .6 compared to 1.5. Output, however, increases by 1, as in the full information case, because capital is fixed. It follows that consumption has to pick up the difference by increasing to 1.2, compared to .8. We know that the key to sustaining confounding dynamics is in the behavior of the interest rate, r_t. The interest rate increases at impact by 1, mirroring productivity, while it drops only to around .5 in the subsequent period, due to the limited capital investment at impact. Because agents cannot observe aggregate capital, the limited drop in the interest rate "confuses" them in the following sense: in their learning effort, they input a realization for the interest rate in which a past innovation – the unitary jump in productivity in this case – has a larger effect than the current innovation,

Figure 2: Impulse Responses to Aggregate Productivity Shock



Impulse response of Capital (Investment) k_{t+1}, Output y_t, Consumption c_t, and Interest Rate r_t, to a unitary positive shock to aggregate productivity a_t. The dashed-black line represents the response under the full information equilibrium (5.17); the solid-blue line represents the response under the equilibrium with confounding dynamics (5.18). The parameter values are $\beta = 0.985$, $\alpha = 0.33$, $\rho = 1$, $\sigma_v/\sigma_e = 2.5$. The roots of the characteristic polynomial $\Phi(L)$ are $\zeta = 0.32$, and $\tilde{\zeta} = 3.03$. The equilibrium with confounding dynamics is computed using $\lambda = -0.78$.

which is zero. Agents then (optimally) over-correct. The end result is an average belief that, after impact, productivity is higher than previously thought, which triggers an overly-optimistic investment demand with capital topping out at 2.25 in period 1, compared to the climb to 1.9 under full information. Because productivity is not growing after impact, output in period 1 only grows due to capital invested in period 0, which was low. As a consequence, to allow for the large capital investment, consumption has to drop in period 1 to half of its increase at impact. Note that the drop in consumption is optimal, given that the interest rate is higher than the full information counterpart. In period 2, the same learning effort that created an over-reaction to the initial productivity innovation implies that beliefs about the innovation are now pessimistic, which triggers a slow-down in capital demand. However, output is now boosted from the previous period's optimistic capital demand, and so consumption picks up the difference by increasing to 1.8, with the interest rate now markedly dropping into negative territory.

The following periods all mimic the over-under-reactions just described, with a declining magnitude, as the initial innovation to productivity decays. The slower the decay (the larger the value of ρ), the more this economy deviates from the full-information equilibrium. This is why confounding dynamics are more likely with higher values of ρ as reported in Table 1.

We believe that the numerical application of Figure 2 displays a qualitative behavior that is interesting and promising for quantitative applications. First, we are able to generate dynamics that resemble waves of optimism and pessimism, just as a consequence of optimal learning from endogenous variables. Second, while the period of such waves is mechanically determined by the assumption that we look at equilibria with only one non-invertible root λ, richer non-invertible conditions – such as ones with multiple roots, conjugate pairs, etc. – result in waves that can be longer and asymmetric in duration (we explore a simple example with multiple roots in Appendix B.2). Third, we have assumed away additional sources of noise and information, such as exogenous noisy signals, in order to keep things analytically tractable and transparent. However, we envision a richer environment with several types of signals, but where confounding dynamics remain a major determinant of equilibrium behavior.

6 Concluding Comments

We have introduced a rational expectations equilibrium that generates over- and under-reaction relative to its complete information counterpart. This systematic optimism and pessimism is generated from a simple and optimal learning mechanism that can be easily applied to any dynamic setting. Future work will seek to better understand the empirical properties of confounding dynamics by incorporating them into real and nominal business cycle models designed to be taken to data. Theoretical results of Section 5 and preliminary empirical results of Hur, Leeper, Rondiana, and Walker (2016) show much promise. Hur, Leeper, Rondiana, and Walker (2016) apply Bayesian methods to estimate both a real business cycle model and a medium-sized new Keynesian model with and without confounding dynamics. Results suggest that allowing for confounding dynamics substantially increases the empirical fit of these models relative to competing information specifications (e.g., news shocks, complete information). Future work [Huertgen, Hoffmann, Rondina, and Walker (2016)] will also seek to show an equivalence between the analytic function approach advocated here and the more familiar time-domain approach. Contrasting these approaches in a side-by-side fashion will help to highlight the benefits of the analytic function approach while (hopefully) demystifying certain aspects of it.

References

Adam, K. (2007): "Optimal Monetary Policy with Imperfect Common Knowledge," *Journal of Monetary Economics*, 54(2), 267–301.

Allen, F., S. Morris, and H. Shin (2006): "Beauty contests and iterated expectations in asset markets," *Review of Financial Studies*, 19(3), 719–752.

Angeletos, G.-M., and J. La'O (2009): "Noisy Business Cycles," *NBER Macroeconomics Annual*, 24.

Angeletos, G.-M., and J. La'O (2013): "Sentiments," *Econometrica*, 81(2), 739–780.

Angeletos, G.-M., and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75(4).

Bacchetta, P., and E. van Wincoop (2006): "Can Information Heterogeneity Explain the Exchange Rate Puzzle?," *American Economic Review*, 96(3), 552–576.

Benhabib, J., P. Wang, and Y. Wen (2015): "Sentiments and Aggregate Demand Fluctuations," *Econometrica*, 83(2), 549585.

Bernhardt, D., P. Seiler, and B. Taub (2010): "Speculative Dynamics," *Economic Theory*, 44, 1–52.

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–1312.

Brockwell, P. J., and R. A. Davis (1987): *Time Series: Theory and Methods*. Springer Series in Statistics, Springer-Verlag New York.

Brown, J., and R. Churchill (2013): *Complex Variables and Applications*. McGraw-Hill Education, New York, USA, ninth edn.

Cagan, P. D. (1956): "The Monetary Dynamics of Hyperinflation," in *Studies in the Quantity Theory of Money*, ed. by M. Friedman. University of Chicago Press.

Fernandez-Villaverde, J., J. F. Rubio, T. Sargent, and M. Watson (2007): "A,B,C (and D)'s for Understanding VARs," *American Economic Review*, 97, 1021–1026.

Futia, C. A. (1981): "Rational Expectations in Stationary Linear Models," *Econometrica*, 49(1), 171–192.

Galì, J. (2008): *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press, Princeton and Oxford.

Graham, L., and S. Wright (2010): "Information, Heterogeneity and Market Incompleteness," *Journal of Monetary Economics*, 57(2), 164–174.

Gregoir, S., and P. Weill (2007): "Restricted Perception Equilibria and Rational Expectation Equilibrium," *Journal of Economic Dynamics and Control*, 31(1), 81–109.

Grossman, S. J., and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393–408.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7–46.

Hassan, T. A., and T. M. Mertens (2011): "The Social Cost of Near-Rational Investment," NBER Working Paper 17027.

Hellwig, C. (2006): "Monetary Business Cycle Models: Imperfect Information," *New Palgrave Dictionary of Economics*.

Hellwig, C., and V. Venkateswaran (2009): "Setting the Right Prices for the Wrong Reasons," *Journal of Monetary Economics*, 56, S57–S77.

Hoffman, K. (1962): *Banach Spaces of Analytic Functions*. Prentice-Hall, Englewood Cliffs, New Jersey.

Huertgen, P., M. Hoffmann, G. Rondina, and T. B. Walker (2016): "Rational Expectations with Dispersed Information: A Comparison of Solution Techniques," CAEPR Working Paper.

Huo, Z., and N. Takayama (2016): "Rational Expectations Models with Higher Order Beliefs," Working Paper.

Hur, J., E. M. Leeper, G. Rondiana, and T. B. Walker (2016): "No News is Good News," Caepr discussion paper.

Kasa, K. (2000): "Forecasting the Forecasts of Others in the Frequency Domain," *Review of Economic Dynamics*, 3, 726–756.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014): "Heterogenous Beliefs and Tests of Present Value Models," *Review of Economic Studies*, 81(3), 1137–1163.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*. Macmillan, London, U.K.

King, R. (1982): "Monetary Policy and the Information Content of Prices," *Journal of Political Economy*, 90(2), 247–279.

Klein, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

Lorenzoni, G. (2009): "A Theory of Demand Shocks," *American Economic Review*, 99(5), 2050–2084.

Lucas, Jr., R. E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4, 103–124.

——— (1975): "An Equilibrium Model of the Business Cycle," *Journal of Political Economy*, 83, 1113–1144.

Makarov, I., and O. Rytchkov (2012): "Forecasting the Forecasts of Others: Implications for Asset Pricing," *Journal of Economic Theory*, 147, 941–966.

Melosi, L. (2016): "Signaling Effects of Monetary Policy," forthcoming, *Review of Economic Studies*.

Nimark, K. (2008): "Dynamic Pricing and Imperfect Common Knowledge," *Journal of Monetary Economics*, (55), 365–382.

Nimark, K. (2010): "Dynamic higher order expectations," Working paper, Universitat Pompeu Fabra.

Pearlman, J. G., and T. J. Sargent (2005): "Knowing the Forecasts of Others," *Review of Economic Dynamics*, 8(2), 480–497.

Pigou, A. C. (1929): *Industrial Fluctuations*. Macmillan, London, second edn.

Radner, R. (1979): "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 47(3), 655–678.

Rondina, G. (2009): "Incomplete Information and Informative Pricing," Working Paper. UCSD.

Rozanov, Y. A. (1967): *Stationary Random Processes*. Holden-Day, San Francisco.

Sargent, T. (1987): *Macroeconomic Theory*. Emerald Group Publishing Limited, Bingley, UK, second edn.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Singleton, K. J. (1987): "Asset Prices in a Time Series Model with Disparately Informed, Competitive Traders," in *New Approaches to Monetary Economics*, ed. by W. Barnett, and K. Singleton. Cambridge University Press, Cambridge.

Taub, B. (1989): "Aggregate Fluctuations as an Information Transmission Mechanism," *Journal of Economic Dynamics and Control*, 13(1), 113–150.

Townsend, R. M. (1983a): "Equilibrium Theory with Learning and Disparate Expectations: Some Issues and Methods," in *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*, ed. by E. S. Phelps, and R. Frydman, pp. 169–202. Cambridge University Press.

——— (1983b): "Forecasting the Forecasts of Others," *Journal of Political Economy*, 91(4), 546–588.

Uhlig, H. (1999): "Analyzing Non Linear Dynamics Stochastic Models," in *Computational Methods for the Study of Dynamic Economies*, ed. by A. S. R. Marimon, pp. 30–61. Oxford University Press, Oxford.

Walker, T. B. (2007): "How Equilibrium Prices Reveal Information in Time Series Models with Disparately Informed, Competitive Traders," *Journal of Economic Theory*, 137(1), 512–537.

Wang, J. (1993): "A Model of Asset Prices under Asymmetric Information," *Review of Economic Studies*, 60(2), 249–282.

Whiteman, C. (1983): *Linear Rational Expectations Models: A User's Guide*. University of Minnesota Press, Minneapolis.

Whittle, P. (1983): *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press, Minneapolis, second edn.

Woodford, M. (2003a): "Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information, and Expectations in Modern Macroeconomics*, ed. by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford. Princeton University Press, Princeton, N.J.

——— (2003b): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ.

A Proofs

A.1 Equivalence in Signal Extraction We need to show that the representations (2.7) and (2.12) are equivalent in terms of unconditional forecast error variance

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|x^t\right)\right)^2\right] = \mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|z^t\right)\right)^2\right] \tag{A.1}$$

when $\lambda^2 = 1 + \sigma_\eta^2/\sigma_\varepsilon^2$.

The optimal forecast $\mathbb{E}[\varepsilon_t|z^t]$ is given by weighting z_t according to the relative variance of ε, $\mathbb{E}(\varepsilon_t|z^t) = \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\eta^2}\right)z_t$ and therefore,

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|z^t\right)\right)^2\right] = \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_\varepsilon^2+\sigma_\eta^2} \tag{A.2}$$

Calculating the optimal expectation for ε_t conditional on x^t requires more careful treatment. While there are many moving average representations for x_t that deliver the same observed autocorrelation structure (which is essentially all the information contained in x^t), there exists only one that minimizes the variance of the forecast error in the LHS of (A.1). We first need to take the conditional expectation $\mathbb{E}[\varepsilon_t|x^t]$. This expectation is found by deriving the fundamental moving-average representation and using the Wiener-Kolmogorov optimal prediction formula. The fundamental representation is derived through the use of Blaschke factors

$$x_t = (L+\lambda)\left(\frac{1+\lambda L}{L+\lambda}\right)\left(\frac{L+\lambda}{1+\lambda L}\right)\varepsilon_t = (1+\lambda L)\tilde{\varepsilon}_t \tag{A.3}$$

where $\tilde{\varepsilon}_t$ is defined as in (2.11). Given that (A.3) is an invertible representation then the Hilbert space spanned by current and past x_t is equivalent to the space spanned by current and past $\tilde{\varepsilon}_t$. This implies

$$\mathbb{E}(\varepsilon_t|\tilde{\varepsilon}^t) = \mathbb{E}(\varepsilon_t|x^t) \tag{A.4}$$

To show (A.4) notice that (A.3) can be written as

$$\varepsilon_t = C(L)\tilde{\varepsilon}_t = \left[\frac{(\lambda^{-1}+L^{-1})}{1-(-\lambda L)^{-1}}\right]\tilde{\varepsilon}_t \tag{A.5}$$

Thus, while (A.3) does not have an invertible representation in current and past $\tilde{\varepsilon}$ it does have a valid expansion in current and future $\tilde{\varepsilon}$. Notice that

$$\varepsilon_t = (\lambda^{-1}+L^{-1})\sum_{j=0}^{\infty}(-\lambda)^{-j}\tilde{\varepsilon}_{t+j} = (\lambda^{-1}+L^{-1})[\tilde{\varepsilon}_t + (-\lambda)^{-1}\tilde{\varepsilon}_{t+1} + \cdots]$$

The optimal prediction formula yields

$$\mathbb{E}(\varepsilon_t|\tilde{\varepsilon}^t) = \left[C(L)\right]_+\tilde{\varepsilon}_t = \lambda^{-1}\tilde{\varepsilon}_t = \left[\frac{1}{\lambda^2(1+\lambda^{-1}L)}\right]x_t \tag{A.6}$$

We must now calculate

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|x^t\right)\right)^2\right] = \mathbb{E}\left(\varepsilon_t^2\right) + \mathbb{E}\left(\varepsilon_t|x^t\right)^2 - 2\mathbb{E}\left(\varepsilon_t\mathbb{E}\left(\varepsilon_t|x^t\right)\right) \tag{A.7}$$

$$= \sigma_\varepsilon^2 + \frac{1}{\lambda^2}\mathbb{E}(\tilde{\varepsilon}_t^2) - \frac{2}{\lambda}\mathbb{E}(\varepsilon_t\tilde{\varepsilon}_t) \tag{A.8}$$

Notice that the squared modulo of the Blaschke factor is equal to 1, $\left(\frac{1+\lambda z}{z+\lambda}\right)\left(\frac{1+\lambda z^{-1}}{z^{-1}+\lambda}\right) = 1$, and therefore $\mathbb{E}(\tilde{\varepsilon}^2) = \sigma_\varepsilon^2$.

To calculate $\mathbb{E}(\varepsilon_t\tilde{\varepsilon}_t)$ we use complex integration and the theory of the residue calculus,

$$\mathbb{E}(\varepsilon_t\tilde{\varepsilon}_t) = \frac{\sigma_\varepsilon^2}{2\pi i}\oint \frac{1+\lambda z}{z+\lambda}\frac{dz}{z} = \sigma_\varepsilon^2\left[\lim_{z\to 0}\frac{1+\lambda z}{z+\lambda}\right] = \frac{\sigma_\varepsilon^2}{\lambda}. \tag{A.9}$$

Equations (A.8) and (A.9) give the desired result

$$\mathbb{E}\left[\left(\varepsilon_t - E\left(\varepsilon_t|x^t\right)\right)^2\right] = \left(1-\frac{1}{\lambda^2}\right)\sigma_\varepsilon^2$$

A.2 Full Information Solution This section derives the solution under full information for model (3.1)-(3.4). The fixed point condition under full information can be found by substituting (4.4)-(4.6) into (3.1), and using (A.14) to substitute for terms featuring $\mathcal{X}(L)$, so that

$$\begin{aligned}&\phi_x\big[\mathcal{X}(L)-\mathcal{X}(0)\big]L^{-1}\varepsilon_t + \phi_x\big[V(L)-V(0)\big]L^{-1}v_{it} + \phi_y\big[\mathcal{Y}(L)-\mathcal{Y}(0)\big]L^{-1}\varepsilon_t + \phi_\theta\big[A(L)-A(0)\big]L^{-1}\varepsilon_t\\ &= \psi_x(L)\mathcal{X}(L)\varepsilon_t + \psi_x(L)\mathcal{V}(L)v_{it} + \psi_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}.\end{aligned} \tag{A.10}$$

This equation defines a fixed point condition for $\mathcal{V}(L)$ with all the terms that multiply v_{it}. Collecting terms that multiply v_{it}, multiplying both sides by L and rearranging we get

$$\mathcal{V}(L)\big(\phi_x + \psi_x(L)L\big) = \phi_x\mathcal{V}(0) + \psi_\theta(L)L. \tag{A.11}$$

Note that $\phi_x(L) \equiv \phi_x + \psi_x(L)L$, which, under assumption (s), has exactly one zero inside the unit circle, denoted by ζ_x. We thus pick $\mathcal{V}(0)$ to remove such zero by setting

$$\phi_x\mathcal{V}(0) + \psi_\theta(\zeta_x)\zeta_x = 0. \tag{A.12}$$

Solving for $\mathcal{V}(0)$, substituting back into (A.11) one finally obtains

$$\mathcal{V}(L) = \frac{\psi_\theta(L)L - \psi_\theta(\zeta_x)\zeta_x}{\phi_x(L)}. \tag{A.13}$$

We now focus on the fixed point for $\mathcal{Y}(L)$ and $\mathcal{X}(L)$. As remarked in the text, the fixed point condition does not feature any components of $\mathcal{V}(L)$, so that one does not need to solve for the latter to obtain the former. To proceed with the solution there are two possibilities: solve for $\mathcal{Y}(L)$ and then recover $\mathcal{X}(L)$, or viceversa. In general, both routes are possible, but there are situations in which one direction is substantially easier than the other. This depends on whether $\gamma_x(0) \neq 0$ or $\gamma_y(0) \neq 0$. We report here both cases. We first consider the case that works whenever $\gamma_x(0) \neq 0$. We begin by manipulating condition (3.4) to get the following relationship between $\mathcal{X}(L)$ and $\mathcal{Y}(L)$,

$$\mathcal{X}(L) = \tilde{\gamma}_y(L)\mathcal{Y}(L) + \tilde{\gamma}_\theta(L)A(L), \tag{A.14}$$

where $\tilde{\gamma}_y(L) = -\frac{\gamma_y(L)}{\gamma_x(L)}$, and $\tilde{\gamma}_\theta(L) = -\frac{\gamma_\theta(L)}{\gamma_x(L)}$. Using (A.14) to substitute for terms featuring $\mathcal{X}(L)$ in (A.10) one obtains

$$\begin{aligned}&\phi_x\big[\tilde{\gamma}_y(L)\mathcal{Y}(L) - \tilde{\gamma}_y(0)\mathcal{Y}(0)\big]L^{-1}\varepsilon_t + \phi_x\big[\tilde{\gamma}_\theta(L)A(L) - \tilde{\gamma}_\theta(0)A(0)\big]L^{-1}\varepsilon_t + \phi_x\big[V(L)-V(0)\big]L^{-1}v_{it}\\ &+ \phi_y\big[\mathcal{Y}(L)-\mathcal{Y}(0)\big]L^{-1}\varepsilon_t + \phi_\theta\big[A(L)-A(0)\big]L^{-1}\varepsilon_t = \psi_x(L)\tilde{\gamma}_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_x(L)\tilde{\gamma}_\theta(L)A(L)\varepsilon_t\\ &+ \psi_x(L)\mathcal{V}(L)v_{it} + \psi_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}.\end{aligned} \tag{A.15}$$

Taking all the terms that multiply ε_t in (A.15), multiplying by L both sides and rearranging, one gets

$$\mathcal{Y}(L)\Phi(L) = \mathcal{Y}(0)\big(\phi_x\tilde{\gamma}_y(0) + \phi_y\big) - \xi_y(L), \tag{A.16}$$

where

$$\xi_y(L) \equiv \big(\phi_x - \psi_x(L)L\big)\tilde{\gamma}_\theta(L)A(L) + \big(\phi_\theta - \psi_\theta(L)L\big)A(L) - \big(\phi_x\tilde{\gamma}_\theta(0) + \phi_\theta\big)A(0). \tag{A.17}$$

Under assumption $(\mathcal{S})$, $\Phi(L)$ has exactly one zero inside the unit circle, denoted by ζ, which means that we can choose $\mathcal{Y}(0)$ to remove such zero. We thus set

$$\mathcal{Y}(0)\big(\phi_x\tilde{\gamma}_y(0)+\phi_y\big)-\xi_y(\zeta)=0. \tag{A.18}$$

Solving for $\mathcal{Y}(0)$, substituting into (A.16) and rearranging, one finally gets

$$\mathcal{Y}(L)=\frac{\xi_y(\zeta)-\xi_y(L)}{\Phi(L)}. \tag{A.19}$$

The expression for $\mathcal{X}(L)$ can then be recovered using (A.14). Next we consider the case that works whenever $\gamma_y(0)\neq 0$. We begin by manipulating condition (3.4) to get the following relationship between $\mathcal{X}(L)$ and $\mathcal{Y}(L)$,

$$\mathcal{Y}(L)=\hat{\gamma}_x(L)\mathcal{X}(L)+\hat{\gamma}_\theta(L)A(L), \tag{A.20}$$

where $\hat{\gamma}_x(L)=-\frac{\gamma_x(L)}{\gamma_y(L)}$, and $\hat{\gamma}_\theta(L)=-\frac{\gamma_\theta(L)}{\gamma_y(L)}$. Using (A.20) to substitute for terms featuring $\mathcal{Y}(L)$ in (A.10) one obtains

$$\begin{aligned}
&\phi_x\big[\mathcal{X}(L)-\mathcal{X}(0)\big]L^{-1}\varepsilon_t+\phi_x\big[V(L)-V(0)\big]L^{-1}v_{it}+\phi_y\big[\hat{\gamma}_x(L)\mathcal{X}(L)-\hat{\gamma}_x(L)\mathcal{X}(0)\big]L^{-1}\varepsilon_t\\
&+\phi_y\big[\hat{\gamma}_\theta(L)A(L)-\hat{\gamma}_\theta(L)A(0)\big]L^{-1}\varepsilon_t+\phi_\theta\big[A(L)-A(0)\big]L^{-1}\varepsilon_t=\psi_x(L)\mathcal{X}(L)\varepsilon_t+\psi_x(L)\mathcal{V}(L)v_{it}+\psi_y(L)\hat{\gamma}_x(L)\mathcal{X}(L)\varepsilon_t\\
&+\psi_y(L)\hat{\gamma}_\theta(L)A(L)\varepsilon_t+\psi_\theta(L)A(L)\varepsilon_t+\psi_\theta(L)v_{it}.
\end{aligned} \tag{A.21}$$

Taking all the terms that multiply ε_t in (A.21), multiplying by L both sides and rearranging, one gets

$$\mathcal{X}(L)\Phi_x(L)=\mathcal{X}(0)\big(\phi_x+\phi_y\hat{\gamma}_y(0)\big)-\xi_x(L), \tag{A.22}$$

where

$$\Phi_x(L)=\phi_x+\phi_y\hat{\gamma}_x(L)-\psi_x(L)L-\psi_y(L)\hat{\gamma}_x(L)L, \tag{A.23}$$

and

$$\xi_x(L)\equiv\big(\phi_y-\psi_y(L)L\big)\hat{\gamma}_\theta(L)A(L)+\big(\phi_\theta-\psi_\theta(L)L\big)A(L)-\big(\phi_y\hat{\gamma}_\theta(0)+\phi_\theta\big)A(0). \tag{A.24}$$

Analogously to Assumption $(\mathcal{S})$, let us assume that $\Phi_x(L)$ has exactly one zero inside the unit circle, denoted by $\hat{\zeta}$, which means that we can choose $\mathcal{X}(0)$ to remove such zero. We thus set

$$\mathcal{X}(0)\big(\phi_x+\phi_y\hat{\gamma}_y(0)\big)-\xi_x(\hat{\zeta})=0. \tag{A.25}$$

Solving for $\mathcal{X}(0)$, substituting into (A.22) and rearranging, one finally gets

$$\mathcal{X}(L)=\frac{\xi_x(\hat{\zeta})-\xi_x(L)}{\Phi_x(L)}. \tag{A.26}$$

The expression for $\mathcal{Y}(L)$ can then be recovered using (A.20).

A.3 Proof of Theorem 1

Factorization We operationalize the key requirement that $Y(\lambda)=0$ for $\lambda\in(-1,1)$ by specifying a guess of the form $Y(L)=(L-\lambda)G(L)$, where $G(L)$ has no zeros inside the unit circle. The first step in the proof is to then use the equilibrium guess to derive the canonical factorization for the information set, so that the Wiener-Kolmogorov formula (2.4) can be applied. The information set can be written as

$$\begin{pmatrix}\theta_{it}\\ y_t\end{pmatrix}=\begin{bmatrix}A(L)\sigma_\varepsilon & \sigma_v\\ (L-\lambda)G(L)\sigma_\varepsilon & 0\end{bmatrix}\begin{pmatrix}\tilde{\varepsilon}_t\\ \tilde{v}_{it}\end{pmatrix}. \tag{A.27}$$

where $\varepsilon_t = \sigma_\varepsilon \tilde{\varepsilon_t}$, $v_{it} = \sigma_v \tilde{v_{it}}$, is a convenient normalization so that the variance-covariance matrix of the innovations vector is the identity matrix. It follows that

$$\Gamma(L) = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ (L-\lambda)G(L)\sigma_\varepsilon & 0 \end{bmatrix}. \tag{A.28}$$

The following Lemma shows the canonical factorization for $\Gamma(L)$.

Lemma A1. *The canonical factorization $\Gamma^*(z)\Gamma^*(z^{-1})^T$ of the variance-covariance matrix $\Gamma(z)\Gamma(z^{-1})^T$, where $\Gamma(z)$ is defined in* (A.28)*, is given by*

$$\Gamma^*(z) = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2+\sigma_v^2}} \begin{bmatrix} A(z)A(\lambda)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v \frac{1-\lambda z}{z-\lambda}\big(A(z)-A(\lambda)\big) \\ A(\lambda)\sigma_\varepsilon^2(z-\lambda)G(z) & \sigma_\varepsilon\sigma_v G(z)(1-\lambda z) \end{bmatrix}. \tag{A.29}$$

Proof. Using Rozanov (1967) procedure, $\Gamma^*(z)$ is computed as

$$\Gamma^*(z) = \Gamma(z) W_\lambda B_\lambda(z). \tag{A.30}$$

where

$$W_\lambda = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2+\sigma_v^2}} \begin{bmatrix} A(\lambda)\sigma_\varepsilon & -\sigma_v \\ \sigma_v & A(\lambda)\sigma_\varepsilon \end{bmatrix}, \quad \text{and} \quad B_\lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1-\lambda z}{z-\lambda} \end{bmatrix}. \tag{A.31}$$

The form of W_λ is obtained by application of Lemma C1 in the Online Technical Appendix. Solving out the matrix multiplication after some algebra one obtains (A.29). □

Expectations Equipped with the canonical factorization (A.29), we next derive the three expectational terms: $\mathbb{E}_{it}(x_{it+1})$, $\mathbb{E}_{it}(y_{t+1})$, and $\mathbb{E}_{it}(\theta_{it+1})$ (recall that $\mathbb{E}_{it}(\theta_{it+1}) = \mathbb{E}_{it}(\theta_{t+1})$). The second and third in the list are given by

$$\mathbb{E}_{it}\begin{pmatrix} \theta_{it+1} \\ y_{t+1} \end{pmatrix} = \left[L^{-1}\Gamma^*(L)\right]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix}. \tag{A.32}$$

Recalling that $\left[L^{-1}\Gamma^*(L)\right]_+ = \left[\Gamma^*(L) - \Gamma^*(0)\right]L^{-1}$, and defining $\tau = \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2+\sigma_v^2}$ one gets

$$\mathbb{E}_{it}(\theta_{t+1}) = \left[A(L) - A(0)\right]L^{-1}\varepsilon_t - (1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\left[A(\lambda) - A(0)\right]\varepsilon_t - \tau\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\left[1 - \tfrac{A(0)}{A(\lambda)}\right]v_{it}, \tag{A.33}$$

$$\mathbb{E}_{it}(y_{t+1}) = \left[(L-\lambda)G(L) + \lambda G(0)\right]L^{-1}\varepsilon_t - (1-\tau)\tfrac{1-\lambda^2}{(1-\lambda L)}G(0)\varepsilon_t + \tau\tfrac{1-\lambda^2}{(1-\lambda L)}\tfrac{G(0)}{A(\lambda)}v_{it}. \tag{A.34}$$

The term $\mathbb{E}_{it}(x_{it+1})$, is substantially more involved to derive, due to the fact that the correlation between x_{it+1} and θ_{it} exists not only because they both depend on ε_t, but they also both depend on v_{it}. Formally, the application of the Wiener-Kolmogorov formula leads to

$$\mathbb{E}_{it}(x_{it+1}) = \left[L^{-1}g_{x_i,(\theta_i,y)}(L)\big(\Gamma^*(L^{-1})^T\big)^{-1}\right]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix}, \tag{A.35}$$

where $g_{x_i,(\theta_i,y)}(L)$ is the variance-covariance generating function between x_i and the information set. Given the equilibrium guess, such function takes the form

$$g_{x_i,(\theta_i,y)}(L) = \begin{bmatrix} X(L)A(L^{-1})\sigma_\varepsilon^2 + V(L)\sigma_v^2 & X(L)(L^{-1}-\lambda)G(L^{-1})\sigma_\varepsilon^2 \end{bmatrix}. \tag{A.36}$$

It follows that

$$L^{-1}g_{x_i,(\theta_i,y)}(L)\big(\Gamma^*(L^{-1})^T\big)^{-1} = \Big[L^{-1}\big(V(L)\sigma_v^2 + X(L)\sigma_\varepsilon^2 A(\lambda)\big) \quad \sigma_\varepsilon\sigma_v L^{-1}\tfrac{1-\lambda L}{L-\lambda}\Big(X(L) - V(L)A(\lambda)\Big)\Big]. \tag{A.37}$$

The application of the annihilator operator requires to take the annihiland minus the principal part of its Laurent series expansion. All the terms have the usual principal part around $L = 0$. However, the term containing $\frac{1-\lambda L}{L-\lambda}$ also has a principal part around $L = \lambda$, it follows that

$$\begin{aligned}&\left[\left(\tfrac{1-\lambda L}{L-\lambda}\right)\tfrac{1}{L}\big(X(L)-V(L)A(\lambda)\big)\right]_+\\&=L^{-1}\left[\left(\tfrac{1-\lambda L}{L-\lambda}\right)\big(X(L)-V(L)A(\lambda)\big)+\tfrac{1}{\lambda}\big(X(0)-V(0)A(\lambda)\big)\right]-\tfrac{1-\lambda^2}{L-\lambda}\tfrac{1}{\lambda}\big(X(\lambda)-V(\lambda)A(\lambda)\big).\end{aligned} \tag{A.38}$$

Finally one gets

$$\begin{aligned}\mathbb{E}_{it}(x_{it+1}) =& L^{-1}\Big[X(L)-X(0)\Big]\varepsilon_t-(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]\varepsilon_t\\&+L^{-1}\Big[V(L)-V(0)\Big]v_{it}+\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]v_{it}.\end{aligned} \tag{A.39}$$

Fixed Point We begin by manipulating condition (3.4) to get the following relationship between $X(L)$ and $Y(L)$,

$$X(L) = \tilde{\gamma}_y(L)Y(L)+\tilde{\gamma}_\theta(L)A(L), \tag{A.40}$$

where $\tilde{\gamma}_y(L) = -\frac{\gamma_y(L)}{\gamma_x(L)}$, and $\tilde{\gamma}_\theta(L) = -\frac{\gamma_\theta(L)}{\gamma_x(L)}$. Next we substitute the equilibrium guess and expressions (A.33), (A.34), and (A.39) into model (3.1), which leads to the expression

$$\begin{aligned}&\phi_x\Big[L^{-1}\Big[X(L)-X(0)\Big]\varepsilon_t-(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]\varepsilon_t\\&+L^{-1}\Big[V(L)-V(0)\Big]v_{it}+\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]v_{it}\Big]\\&\phi_y\left[\big[(L-\lambda)G(L)+\lambda G(0)\big]L^{-1}\varepsilon_t-(1-\tau)\tfrac{1-\lambda^2}{(1-\lambda L)}G(0)\varepsilon_t+\tau\tfrac{1-\lambda^2}{(1-\lambda L)}\tfrac{G(0)}{A(\lambda)}v_{it}\right]\\&\phi_\theta\left[\big[A(L)-A(0)\big]L^{-1}\varepsilon_t-(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\big[A(\lambda)-A(0)\big]\varepsilon_t-\tau\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\left[1-\tfrac{A(0)}{A(\lambda)}\right]v_{it}\right]\\&=\psi_x(L)\big(X(L)\varepsilon_t+V(L)v_{it}\big)+\psi_y(L)(L-\lambda)G(L)\varepsilon_t+\psi_\theta(L)A(L)\varepsilon_t+\psi_\theta(L)v_{it}.\end{aligned} \tag{A.41}$$

As one would expect, both on the left and right hand sides there are lag polynomials that multiply ε_t and v_{it}. Because the two stochastic process are uncorrelated, the equality must hold independently for the terms that multiply ε_t for those that multiply v_{it}. Taking into account relationship (A.40), equation (A.41) thus defines two fixed points: one for $(L-\lambda)G(L)$ and one for $V(L)$. Differently from the full information case, the fixed point for the aggregate y_t (that defined by the terms multiplying ε_t) also contains elements of the function $V(L)$, more precisely the constant $V(0)-V(\lambda)$. Therefore, in order to solve for $(L-\lambda)G(L)$, we need first to solve for $V(L)$. Taking the fixed point condition for the terms that multiply v_{it}, multiplying both sides by L and rearranging one obtains

$$\begin{aligned}V(L)\phi_x(L) =& \phi_x V(0)-\phi_x\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\left[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\right]L\\&-\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\left[\phi_y G(0)+\phi_\theta\big(A(\lambda)-A(0)\big)\right]L+\psi_\theta(L)L.\end{aligned} \tag{A.42}$$

where $\phi_x(L)\equiv\phi_x-\psi_x(L)L$. Similarly, the fixed point for $(L-\lambda)G(L)$ is

$$\begin{aligned}&(L-\lambda)G(L)\Phi(L) =\\&-\phi_x\tilde{\gamma}_y(0)\lambda G(0)-\phi_x\big(\tilde{\gamma}_\theta(L)A(L)-\tilde{\gamma}_\theta(0)A(0)\big)+\phi_x(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[X(\lambda)-X(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]L\\&+\phi_y\left[\lambda-(1-\tau)\tfrac{1-\lambda^2}{(1-\lambda L)}L\right]G(0)-\phi_\theta\left[\big[A(L)-A(0)\big]-(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\big[A(\lambda)-A(0)\big]L\right]\\&+\psi_x(L)\tilde{\gamma}_\theta(L)A(L)L+\psi_\theta(L)A(L)L.\end{aligned} \tag{A.43}$$

where we have used (A.40) to substitute for, $X(L)-X(0)$, and, $X(L)$, and, $\Phi(L)\equiv\phi_x(L)+\phi_y-\psi_y(L)L$. The next Lemma will prove very useful.

Lemma A2. $V(\lambda)=\tilde{\gamma}_\theta(\lambda)$.

Proof. Evaluate (A.42) at λ and rearrange to obtain

$$V(\lambda)\psi_x(\lambda)\lambda =$$
$$-\phi_x \tfrac{\tau}{A(\lambda)}\left[X(\lambda)-X(0)\right]-\phi_x(1-\tau)\big(V(\lambda)-V(0)\big)-\tfrac{\tau}{A(\lambda)}\left[\phi_y G(0)\lambda+\phi_\theta\big(A(\lambda)-A(0)\big)\right]+\psi_\theta(\lambda)\lambda. \quad \text{(A.44)}$$

Next, evaluate (A.43) at λ and rearrange to obtain

$$\begin{aligned}0 = &-\tau\phi_x\big(X(\lambda)-X(0)\big)+-\phi_x(1-\tau)\big(V(\lambda)-V(0)\big)A(\lambda)-\phi_y\tau G(0)\lambda-\phi_\theta\big(A(\lambda)-A(0)\big)\tau\\ &+\psi_x(\lambda)\tilde{\gamma}_\theta(\lambda)A(\lambda)\lambda+\psi_\theta(\lambda)A(\lambda)\lambda.\end{aligned} \quad \text{(A.45)}$$

Clearly, for (A.44) and (A.45) to hold, assuming $A(\lambda)\neq 0$, $\psi_x(\lambda)\neq 0$ and $\lambda\neq 0$, it must be that $V(\lambda)=\tilde{\gamma}_\theta(\lambda)$. □

We can now use Lemma A2 to substitute for $V(\lambda)$ in (A.42) and (A.43). It follows that to solve for $(L-\lambda)G(L)$ we just need an expression for $V(0)$, to which we now turn. From assumption (s) we know that there is a root ζ_V that needs to be removed for $V(L)$ to be stationary. We achieve this by choosing the appropriate constant $V(0)$ so that the numerator on the right hand side of (A.42) vanishes when evaluated at ζ_V,

$$\begin{aligned}&\phi_x V(0)-\phi_x\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda\zeta_V)}\left[X(\lambda)-X(0)-\big(\tilde{\gamma}_\theta(\lambda)-V(0)\big)A(\lambda)\right]\zeta_V\\ &-\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda\zeta_V)}\left[\phi_y G(0)+\phi_\theta\big(A(\lambda)-A(0)\big)\right]\zeta_V+\psi_\theta(\zeta_V)\zeta_V=0\end{aligned} \quad \text{(A.46)}$$

Using (A.40) so substitute for $X(\lambda)-X(0)$, and rearranging one obtain the expression

$$\phi_x V(0)A(\lambda)=m(\lambda)\big(\phi_x\tilde{\gamma}_y(0)+\phi_y\big)\lambda G(0)+n(\lambda), \quad \text{(A.47)}$$

where

$$m(\lambda)\equiv\frac{\tau(1-\lambda^2)\zeta_V}{(1-\lambda\zeta_V)\lambda-\tau(1-\lambda^2)\zeta_V}, \quad \text{(A.48)}$$

and

$$n(\lambda)\equiv\frac{\phi_\theta\tau(1-\lambda^2)\big(A(\lambda)-A(0)\big)\zeta_V-\phi_x\tau(1-\lambda^2)\tilde{\gamma}_\theta(0)A(0)\zeta_V-\psi_\theta(\zeta_V)\zeta_V\lambda A(\lambda)}{(1-\lambda\zeta_V)\lambda-\tau(1-\lambda^2)\zeta_V}. \quad \text{(A.49)}$$

Next we used (A.47) in (A.43), and we also substitute $X(\lambda)-X(0)$ using (A.40) to get

$$(L-\lambda)G(L)=\frac{-\lambda G(0)\big(\phi_x\tilde{\gamma}_y(0)+\phi_y\big)H(L)+J(L)}{\Phi(L)(1-\lambda L)\lambda}, \quad \text{(A.50)}$$

where

$$H(L)=\lambda(1-\lambda L)-(1-\tau)(1-\lambda^2)(1+m(\lambda))L, \quad \text{(A.51)}$$

and

$$\begin{aligned}J(L)=&(1-\tau)(1-\lambda^2)\big[n(\lambda)-\phi_x\tilde{\gamma}_\theta(0)A(0)+\phi_\theta\big(A(\lambda)-A(0)\big)\big]L+A(0)\big(\phi_x\tilde{\gamma}_y(0)+\phi_y\big)\lambda(1-\lambda L)\\ &-\big[\big(\phi_x-\psi_x(L)L\big)\tilde{\gamma}_\theta(L)+\phi_\theta-\psi_\theta(L)L\big]A(L)\lambda(1-\lambda L).\end{aligned} \quad \text{(A.52)}$$

Under assumption $(\mathcal{S})$, $\Phi(L)$ has a zero inside the unit circle at ζ, which means that we need to choose the constant $G(0)$ so to cancel it. This is achieved by setting

$$-\lambda G(0)\big(\phi_x\tilde{\gamma}_y(0)+\phi_y\big)H(\zeta)+J(\zeta)=0. \quad \text{(A.53)}$$

Solving for $G(0)$ and substituting back into (A.50) one gets

$$(L-\lambda)G(L)=\frac{J(L)H(\zeta)-J(\zeta)H(L)}{\Phi(L)(1-\lambda L)\lambda}. \quad \text{(A.54)}$$

Next, recall that we defined

$$\xi_y(L) \equiv A(0)\big(\phi_x\tilde{\gamma}_y(0) + \phi_y\big) - \big[\big(\phi_x - \psi_x(L)L\big)\tilde{\gamma}_\theta(L) + \phi_\theta - \psi_\theta(L)L\big]A(L), \tag{A.55}$$

and letting

$$\tilde{\xi} \equiv n(\lambda) - \phi_x\tilde{\gamma}_\theta(0)A(0) + \phi_\theta\big(A(\lambda) - A(0)\big), \tag{A.56}$$

one can show that (A.54) can be written as

$$(L-\lambda)G(L) = \frac{\xi_y(\zeta) - \xi_y(L)}{\Phi(L)} - (1-\tau)(1-\lambda^2)(\zeta - L)\frac{\tilde{\xi} - (1+m(\lambda)\xi_y(\zeta)}{H(\zeta)\Phi(L)(1-\lambda L)}. \tag{A.57}$$

Using the factorization $\Phi(L) = (\zeta - L)\tilde{\Phi}(L)$, and defining

$$\mathcal{A}(\lambda) \equiv \frac{\tilde{\xi} - \big(1+m(\lambda)\big)\xi_y(\zeta)}{H(\zeta)}, \tag{A.58}$$

expression (4.15) follows. Finally, for the solution to be consistent with the information set that we have used to derive it, it must be that the polynomial in (4.15) vanishes at $L = \lambda$, which corresponds to condition (4.16) in the Theorem. We have proven that *if* exactly one $\lambda \in (-1,1)$ exists, then the REE is given by (4.15). For the *only if* part, we need to consider two possibilities: (a) there is no $\lambda \in (-1,1)$ that satisfies condition (4.16); (b) there is more than one $\lambda \in (-1,1)$ that satisfies condition (4.16). In both cases, the information conveyed by y_t in equilibrium is inconsistent with the information used to derive the expectations that we use to determine the fixed point. Hence, in both cases the fixed point is not obtained. More precisely, if (a) is the case, then y_t would be revealing perfectly the history of ε_t, and so the equilibrium will correspond to the full information one, $\mathcal{Y}(L)$. If (b) is the case, then the factorization of $\Gamma(L)$ would be incorrect, as $\Gamma^*(L)$ in (A.29) would still be non-invertible. To see this, consider the example where there are two solutions to (4.16), λ_1 and λ_2, both inside the unit circle. Then, the actual form of the equilibrium process should be $G(L)(L-\lambda_1)(L-\lambda_2)$, but the factorization above only takes care of one root, say λ_1, without loss of generality. It follows that

$$\Gamma^*(L) = \frac{1}{\sqrt{A(\lambda_1)^2\sigma_\varepsilon^2 + \sigma_v^2}}\begin{bmatrix} A(L)A(\lambda_1)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v\frac{1-\lambda_1 L}{L-\lambda_1}\big(A(L) - A(\lambda_1)\big) \\ A(\lambda_1)\sigma_\varepsilon^2(L-\lambda_1)(L-\lambda_2)G(L) & \sigma_\varepsilon\sigma_v G(L)(L-\lambda_2)(1-\lambda_1 L)\end{bmatrix}, \tag{A.59}$$

whose determinant still vanishes at $L = \lambda_2$, so that $\Gamma^*(L)$ is not the appropriate factorization. If (b) is the case one can modify the initial guess and consider $N > 1$ roots inside the unit circle, looking then for a condition like (4.16) to deliver exactly N solutions. We restrict our attention to $N = 1$ for simplicity and because the full description of the space of REE with confounding dynamics is beyond the scope of this paper, but we hope it is clear that our methods extend to the more general case.

No Information from the Model The last thing to check to complete the proof is to ensure that there is no information that is transmitted by a clever manipulation of the model conditions – which are part of the information set of the agents – combined with the knowledge of the history of θ_{it} and y_t. For instance, suppose that the market clearing condition (3.4) is specified so that $\int_0^1 x_{it}\mu(i)di = y_t$, which means that y_t is the aggregate of x_{it}, then this would imply $X(L) = Y(L)$, which would result in $x_{it} - y_t = V(L)v_{it}$. Because rational agents know all this, they know that the difference $x_{it} - y_t$ is just a linear combination of the individual innovations v_{it}. It follows that they could, in principle, back out the realizations of v_{it}'s by inverting $V(L)$. More generally, the link between $X(L)$ and $Y(L)$ due to (3.4) can be used by rational agents to obtain additional information on the underlying innovations. For this not to happen, if one augments the information set of the agents by $x_{it} - y_t$, the information matrix must still be non-invertible at λ. The following Lemma shows that this is indeed the case for the equilibrium of Theorem 1.

Lemma A3. *In the equilibrium with confounding dynamics of Theorem 1, consider the augmented information matrix*

$\tilde{\Gamma}(L)$, *where*

$$\begin{pmatrix} \theta_{it} \\ y_t \\ x_{it} - y_t \end{pmatrix} = \tilde{\Gamma}(L) \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix} = \begin{bmatrix} A(L) & 1 \\ Y(L) & 0 \\ X(L) - Y(L) & V(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix}. \tag{A.60}$$

The 2*-by-*2 *minors of* $\tilde{\Gamma}(L)$ *all vanish at* λ.

Proof. Matrix $\tilde{\Gamma}(L)$ has three minors, whose determinants are, respectively, $Y(L)$, $Y(L)V(L)$, and, $A(L)V(L) - (X(L) - Y(L))$. The first two minors clearly vanish at λ since, by construction, $Y(\lambda) = 0$. For the third minor, use (A.40) to write

$$A(L)V(L) - (X(L) - Y(L)) = A(L)V(L) - \tilde{\gamma}_y(L)Y(L) - \tilde{\gamma}_\theta(L)A(L) + Y(L). \tag{A.61}$$

We thus need to show that

$$A(\lambda)V(\lambda) = \tilde{\gamma}_\theta(\lambda)A(\lambda), \tag{A.62}$$

but this follows immediately from Lemma A2. □

A.4 Proof of Proposition 1 The proof of the result follows the same steps as that of Theorem 1, with the difference that we solve for $X(L)$ first – $K(L)$ in the application. Recall that

$$\phi_x = \alpha\beta, \quad \phi_y = 1 - \alpha\beta, \quad \phi_\theta = 1, \quad \psi_x(L) = \alpha(1+\beta) - \alpha L, \quad \psi_y(L) = 0, \quad \psi_\theta(L) = -1.$$

and

$$\gamma_x(L) = (1-\alpha)L, \quad \gamma_y(L) = 1, \quad \gamma_\theta(L) = -1.$$

Note that, although the notation adopted in the model has the two variables having different time subscripts, r_t and k_{t+1}, they are both pre-determined at time t, and so they are both functions of possibly the infinite history of ε_t up to time t. Since we are looking for an equilibrium with confounding dynamics, we operationalize the condition $R(\lambda) = 0$ by conjecturing

$$R(L) = (L - \lambda)G(L), \tag{A.63}$$

where $G(L)$ has no zeros inside the unit circle. Because in equilibrium $R(L) = A(L) - (1-\alpha)K(L)L$, the conjecture immediately implies

$$A(\lambda) = (1-\alpha)K(\lambda)\lambda, \tag{A.64}$$

a relationship that will be useful in what follows. One important remark on (A.64) is that it implies $\lambda \neq 0$. In fact, evaluating the expression at $\lambda = 0$, provided that $K(0)$ is well defined, which must be the case in the solution we want to characterize, gives $A(0) = 0$, which never holds by assumption. Hence, the statement of the Proposition requires $|\lambda| \in (0,1)$. The information set takes the form of (A.27), where $x_{it} = a_{it}$ and $y_t = r_t$, so that $\mathbb{E}_{it}(a_{t+1})$ and $\mathbb{E}_{it}(r_{t+1})$ are provided by (A.33) and (A.34), respectively. For the term $\mathbb{E}_{it}\big(k_{it+2}\big)$ things require some extra steps. We work under the conjecture that

$$k_{it+1} = K(L)\varepsilon_t + V(L)v_{it}, \tag{A.65}$$

Next, we evaluate the variance-covariance generating function between the information set and k_{it+1}, which is

$$g_{k_i,(a_i,r)}(z) = \begin{bmatrix} K(z)A(z^{-1})\sigma_\varepsilon^2 + V(z)\sigma_v^2 & K(z)(z^{-1} - \lambda)G(z^{-1})\sigma_\varepsilon^2 \end{bmatrix}. \tag{A.66}$$

We then use this expression, together with the canonical factorization $\Gamma^*(z)$ in (A.29) in the Wiener-Kolmogorov formula (2.4), and following steps similar to (A.37) and (A.38) to finally get

$$\begin{aligned} \mathbb{E}_{it}(k_{it+2}) = {} & L^{-1}\Big[K(L) - K(0)\Big]\varepsilon_t - (1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[K(0) - K(\lambda) - \big(V(0) - V(\lambda)\big)A(\lambda)\Big]\varepsilon_t \\ & + L^{-1}\Big[V(L) - V(0)\Big]v_{it} - \tau\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[\tfrac{K(0)-K(\lambda)}{A(\lambda)} + \big(V(0) - V(\lambda)\big)\Big]v_{it}. \end{aligned} \tag{A.67}$$

We can now use the expressions for the expectational terms to obtain a fixed point condition similar to (A.41),

$$
\begin{aligned}
&\alpha(1+\beta)K(L)\varepsilon_t + \alpha(1+\beta)V(L)v_{it} = \alpha\beta L^{-1}\Big[K(L)-K(0)\Big]\varepsilon_t + \alpha\beta L^{-1}\Big[V(L)-V(0)\Big]v_{it} \\
&-\alpha\beta(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[K(\lambda)-K(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]\varepsilon_t + \alpha\beta\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[K(\lambda)-K(0)-\big(V(\lambda)-V(0)\big)A(\lambda)\Big]v_{it} \\
&+\alpha K(L)L\varepsilon_t + \alpha V(L)Lv_{it} + A(L)\varepsilon_t + v_{it} - \big[A(L)-A(0)\big]L^{-1}\varepsilon_t + (1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\big[A(\lambda)-A(0)\big]\varepsilon_t \\
&-\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}\Big[A(\lambda)-A(0)\Big]v_{it} + (1-\alpha\beta)\Big[A(L)-(1-\alpha)K(L)L-A(0)\Big]L^{-1}\varepsilon_t \\
&+(1-\alpha\beta)(1-\tau)\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}A(0)\varepsilon_t - (1-\alpha\beta)\tfrac{\tau}{A(\lambda)}\tfrac{1-\lambda^2}{\lambda(1-\lambda L)}A(0)v_{it},
\end{aligned}
\tag{A.68}
$$

where we have used $(L-\lambda)G(L) = A(L)-(1-\alpha)K(L)L$, and thus $-\lambda G(0) = A(0)$, to substitute for terms related to $G(L)$. The fixed point equation contains only terms related to the endogenous polynomials $V(L)$ and $K(L)$, and one can proceed to solve for the fixed point as in the proof of Theorem 1. In particular, using the same steps as in Lemma A2, one can show that $A(\lambda)V(\lambda) = K(\lambda)$, and, in addition, we know that (A.64) holds, so we can set $K(\lambda) = \frac{A(\lambda)}{\lambda(1-\alpha)}$. The uniqueness of a stationary solution under Assumption (s) and condition (5.11), is once again obtained by the appropriate choice of $V(0)$ and $K(0)$. In the end, the expression for $\mathcal{A}_k(\lambda)$, analogue to the constant $\mathcal{A}(\lambda)$ in Theorem 1, can be simplified to

$$
\mathcal{A}_k(\lambda) = \frac{\big[(1-\lambda\beta)\lambda-\tau(1-\lambda^2)\beta\big](1+\beta)A(\lambda)-\alpha\beta^2(1-\tau)(1-\lambda^2)A(0)+(\zeta-\alpha\beta)\lambda(1-\lambda\beta)A(\zeta)}{\alpha(1+\beta)\big[(1-\lambda\beta)\lambda-\tau(1-\lambda^2)(\beta-\zeta)\big]}
\tag{A.69}
$$

The condition for the existence of exactly one $|\lambda| \in (0,1)$ follows from using $K(L)$ to write $R(L)$ and then imposing $R(\lambda) = 0$. The same argument that we have used in the proof of Theorem 1 to argue for the *if and only if* requirements applies here too. This completes the proof.

B Additional Results and Derivations (for Online Publication)

B.1 Derivation of Predictions in Section 2.2 The critical step in obtain Equation (2.9) is the solution to the annihilating operator function

$$
\Big[\sigma_\varepsilon^2(\lambda+L^{-1})\big[(1+\lambda L^{-1})\sigma_\varepsilon\big]^{-1}\Big]_+ = \sigma_\varepsilon\left[\frac{\lambda+L^{-1}}{1+\lambda L^{-1}}\right]_+ .
\tag{B.1}
$$

Multiply both numerator and denominator of the right hand side by L to get

$$
\sigma_\varepsilon\left[\frac{\lambda+L^{-1}}{1+\lambda L^{-1}}\right]_+ = \sigma_\varepsilon\left[\frac{1+\lambda L}{L+\lambda}\right]_+ .
\tag{B.2}
$$

The argument is now a regular function that has an isolated singularity at $L=-\lambda$. The principal part of the Laurent series expansion is determined as

$$
\lim_{z\to-\lambda}(L+\lambda)\frac{1+\lambda L}{L+\lambda} = (1-\lambda^2)
\tag{B.3}
$$

Using Lemma C2 and after some algebra we have

$$
\left[\frac{1+\lambda L}{L+\lambda}\right]_+ = \frac{1+\lambda L}{L+\lambda} - \frac{1-\lambda^2}{L+\lambda} = \frac{\lambda(L+\lambda)}{L+\lambda} = \lambda.
\tag{B.4}
$$

Equation (2.9) immediately follows. In Section B.2 we show that (B.4) generalizes to the case of N singularities.

B.2 N Non-Invertible Roots The following proposition describes the prediction formula for the innovations of a process with N non-invertible roots.

Proposition B1. *Let*

$$s_t = \prod_{i=1}^{N} (L - \lambda_i)\varepsilon_t. \tag{B.5}$$

The least squares prediction $P(\varepsilon_t|s^t)$ *is given by*

$$P(\varepsilon_t|s^t) = \prod_{i=1}^{N} \lambda_i \frac{L - \lambda_i}{1 - \lambda_i L}\varepsilon_t. \tag{B.6}$$

Proof. The first step in the proof is to figure out the canonical factorization of $g_{ss}(z)$ when s_t is as in (B.5). Rozanov (1967) method applies directly here so that

$$\Gamma^*(z) = \sigma_\varepsilon \prod_{i=1}^{N}(1 - \lambda_i L). \tag{B.7}$$

The application of the Wiener-Kolmogorov prediction formula results in the following

$$\Pi_0(z) = \left[\sigma_\varepsilon^2 \prod_{i=1}^{N}(\lambda_i - z^{-1})\big[\sigma_\varepsilon \prod_{i=1}^{N}(1 - \lambda_i z^{-1})\big]^{-1}\right]_+ \big(\prod_{i=1}^{N}(1 - \lambda_i L)\big)^{-1}. \tag{B.8}$$

The next Lemma is useful in solving for the annihilating operator.

Lemma B1.

$$\left[\prod_{i=1}^{N} \frac{\lambda_i - z^{-1}}{1 - \lambda_i z^{-1}}\right]_+ = \prod_{i=1}^{N} \lambda_i. \tag{B.9}$$

The proof of the Lemma is by induction, repeatedly using Lemma C2 to obtain a solution for $N = 1, 2, 3, ...$ □

Application of the Lemma leads to

$$\Pi_0(z) = \prod_{i=1}^{N} \frac{\lambda_i}{1 - \lambda_i z}. \tag{B.10}$$

The final result of the proposition then immediately follows.

Figure 3 shows the impulse response of the prediction formula for $N = 2$ and $N = 3$. As one can see, the mechanical alternation of over-reaction and under-reaction is not present now. Rather, longer and asymmetric cycles are possible.

Elaborate more here

B.3 CONFOUNDING DYNAMICS WITH RECTANGULAR $\Gamma(L)$

Consider a process s_t specified as

$$s_t = \lambda\varepsilon_t + \varepsilon_{t-1} + v_t, \tag{B.11}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon)$ and $v_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_v)$. Here $\Gamma(L)$ is the 2×1 rectangular matrix

$$\Gamma(L) = \begin{pmatrix} (\lambda + L)\sigma_\varepsilon \\ \sigma_v \end{pmatrix}. \tag{B.12}$$

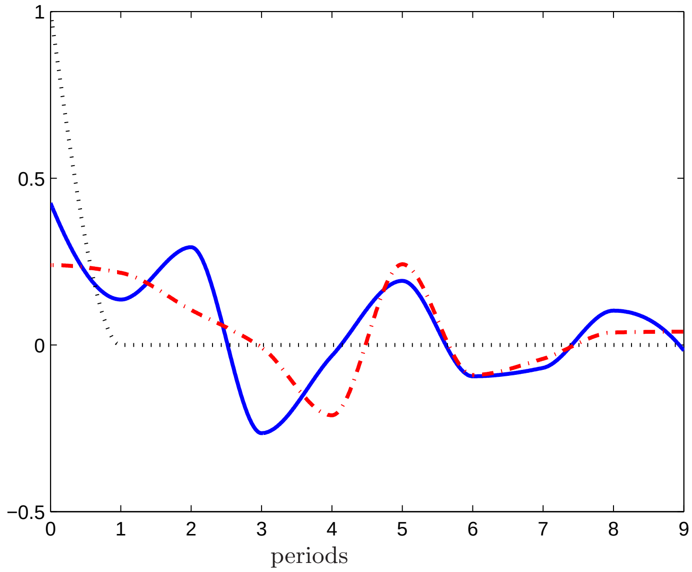
As in Section 2.2, suppose that we want the mean-squared error minimizing prediction for ε_t, given that s^t is observed. The key step is to obtain the canonical factorization of the variance-covariance generating function

$$g_{ss}(z) = \sigma_\varepsilon^2(\lambda + z)(\lambda + z^{-1}) + \sigma_v^2. \tag{B.13}$$

To achieve this we follow the procedure outline in Sargent (1987), Chapter XI. First note that

$$\sigma_\varepsilon^2(\lambda + z)(\lambda + z^{-1}) + \sigma_v^2 = \sigma_\varepsilon^2(1 + z\lambda)(1 + z^{-1}\lambda) + \sigma_v^2. \tag{B.14}$$

Figure 3: Prediction Impulse Response



Impulse response of prediction formula $P(\varepsilon_t|s^t)$), for $N = 2$ (solid blue line), and $N = 3$ (dashed red line). The non-invertible roots are $\lambda_1 = -0.3 + 0.75i$, $\lambda_2 = -0.3 - 0.75i$, and $\lambda_3 = -0.75$.

We are looking for the constants $|\lambda^*| < 1$, and $\sigma_w \geq 0$, so that

$$g_{ss}(z) = \sigma_\varepsilon^2(1 + z\lambda)(1 + z^{-1}\lambda) + \sigma_v^2 = \sigma_w^2(1 + z\lambda^*)(1 + z^{-1}\lambda^*). \tag{B.15}$$

The canonical factorization will then be $\Gamma^*(z) = (1 + \lambda^* z)\sigma_w$. Note that (B.15) can be written as

$$(1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2 + \lambda\sigma_\varepsilon^2 z^{-1} + \lambda\sigma_\varepsilon^2 z = \sigma_w^2(1 + \lambda^{*2}) + \sigma_w^2\lambda^* z^{-1} + \sigma_w^2\lambda^* z. \tag{B.16}$$

Matching coefficients we get two conditions in two unknowns, namely

$$\sigma_w^2\lambda^* = \lambda\sigma_\varepsilon^2, \tag{B.17}$$

and

$$\sigma_w^2(1 + \lambda^{*2}) = (1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2. \tag{B.18}$$

Using (B.17) to substitute for σ_w in (B.18) one gets a quadratic equation for λ^* of the form

$$\lambda\sigma_\varepsilon^2 - \big((1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2\big)\lambda^* + \lambda\sigma_\varepsilon^2\lambda^{*2}. \tag{B.19}$$

To ensure that $\Gamma^*(z)$ is analytic inside the unit circle one needs to choose the root of (B.19) that lies inside the unit circle, so that $|\lambda^*| < 1$. The roots are

$$\lambda^* = \frac{(1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2 \pm \sqrt{\big((1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2\big)^2 - 4\lambda^2\sigma_\varepsilon^4}}{2\lambda\sigma_\varepsilon^2}. \tag{B.20}$$

The choice of the appropriate root depends on whether λ is inside or outside the unit circle. To see this, take the limit of (B.20) for $\sigma_v \to 0$, which corresponds to the case presented in Section 2.2, and note that

$$\lim_{\sigma_v \to 0} \lambda^* \in \{\lambda^{-1}, \lambda\}. \tag{B.21}$$

where the first root is the one associated with the "+" sign for the discriminant term, and the second root the one associated with the "−" sign. Hence, the appropriate root is the first one when $|\lambda| > 1$, and the second one when $|\lambda| < 1$. Expression (B.17) then provides the solution for σ_w. When $\sigma_v^2 > 0$, the same choice of roots applies. It follows that the presence of a non-invertible root in the signal structure (B.11) affects the optimal prediction problem in essentially the same way it affects the prediction problem of Section 2.2. To complete our argument, let $|\lambda| < 1$, then application of the Wiener-Kolmogorov formula leads to

$$P(\varepsilon_t|s^t) = \frac{\lambda^*}{1+\lambda^* L}s_t = \lambda^*\left(\frac{\lambda + L}{1+\lambda^* L}\right)\varepsilon_t + \frac{\lambda^*}{1+\lambda^* L}v_t. \tag{B.22}$$

Expression (B.22) shows that the confounding dynamics hallmark is retained even in the case of a rectangular $\Gamma(L)$. The factor that multiplies ε_t in (B.22) has the same format as the factor in (2.10), with the only difference that the impact is now scaled by λ^*, and the autoregressive root is also λ^*. One important difference in the rectangular case is that the additional noise term also appears in the prediction function, which is represented by the term that multiplies v_t in (B.22). The noise term has a persistent effect on the prediction, with the same autoregressive root as the ε_t term. The take away message from this example is that confounding dynamics can be present in any type of signal structure. The square signal matrix structure that we employ in the text is analytically convenient, and, at the same time, expression (B.22) suggests that it is also without loss of generality, in so far as the purpose is to qualitatively characterize confounding dynamics.

B.4 Derivation of Equation (3.20)

We consider equation (3.1) with $\phi_\theta = 0$, $\varphi_x(L) = 1$, $\varphi_y(L) = 0$, $\varphi_\theta(L) = -1$, and equation (3.4) with $\gamma_x(L) = \gamma_y(L)$, and $\gamma_\theta(L) = 0$, so that

$$x_{it} = \phi_x \mathbb{E}_{it}(x_{it+1}) + \phi_y \mathbb{E}_{it}(y_{t+1}) + \theta_{it} \tag{B.23}$$

and

$$y_t = \int_0^1 x_{it}\mu(i)di. \tag{B.24}$$

Under the form we have assumed for θ_{it}, we always have that for $j \geq 1$,

$$\mathbb{E}_{it+j}(\theta_{it+j+1}) = \mathbb{E}_{it+j}(\theta_{t+j+1}), \tag{B.25}$$

a property that will keep the notation below manageable. To initiate the iterative substitution, take (B.23) one period forward so that

$$x_{it+1} = \phi_x \mathbb{E}_{it+1}(x_{it+2}) + \phi_y \mathbb{E}_{it+1}(y_{t+2}) + \theta_{it+1}. \tag{B.26}$$

Next aggregate (B.26) and apply (B.24) to get

$$y_{t+1} = \phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2}) + \theta_{t+1}. \tag{B.27}$$

Now we can use (B.26) and (B.27) to substitute for x_{it+1} and y_{t+1} in (B.23) to get

$$\begin{aligned} x_{it} = & \quad \phi_x \mathbb{E}_{it}\left[\phi_x \mathbb{E}_{it+1}(x_{it+2}) + \phi_y \mathbb{E}_{it+1}(y_{t+2}) + \theta_{it+1}\right] + \phi_y \mathbb{E}_{it}\left[\phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2}) + \theta_{t+1}\right] + \theta_{it} \\ = & \quad \phi_x\phi_x \mathbb{E}_{it}(x_{it+2}) + \phi_x\phi_y \mathbb{E}_{it}(y_{t+2}) + +\phi_y \mathbb{E}_{it}\left[\phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2})\right] + (\phi_x + \phi_y)\mathbb{E}_{it}(\theta_{t+1}) + \theta_{it} \end{aligned} \tag{B.28}$$

where, in the second line, we applied the Law of Iterated Expectations (LIE), when possible, and we made use of (B.25). Next we want to substitute for x_{it+2} and y_{t+2}, so we carry (B.23) two periods forward so that

$$x_{it+2} = \phi_x \mathbb{E}_{it+2}(x_{it+3}) + \phi_y \mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}. \tag{B.29}$$

aggregate (B.29), and apply (B.24) to get

$$y_{t+2} = \phi_x \bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y \bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}. \tag{B.30}$$

Now we can use (B.29) and (B.30) to substitute for x_{it+2} and y_{t+3} in (B.28) to get

$$
\begin{aligned}
x_{it} &= \phi_x\phi_x\mathbb{E}_{it}\big[\phi_x\mathbb{E}_{it+2}(x_{it+3}) + \phi_y\mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}\big] + \phi_x\phi_y\mathbb{E}_{it}[\phi_x\bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y\bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}] \\
&\quad + \phi_y\mathbb{E}_{it}\big\{\phi_x\bar{\mathbb{E}}_{t+1}\big[\phi_x\mathbb{E}_{it+2}(x_{it+3}) + \phi_y\mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}\big] + \phi_y\bar{\mathbb{E}}_{t+1}[\phi_x\bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y\bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}]\big\} \\
&\quad + (\phi_x+\phi_y)\mathbb{E}_{it}(\theta_{t+1}) + \theta_{it} \\
&= \phi_x\phi_x\phi_x\mathbb{E}_{it}(x_{it+3}) + \phi_x\phi_x\phi_y\mathbb{E}_{it}(y_{t+3}) + \phi_x\phi_y\phi_x\mathbb{E}_{it}\big[\bar{\mathbb{E}}_{t+2}(x_{it+3})\big] + \phi_x\phi_y\phi_y\mathbb{E}_{it}\big[\bar{\mathbb{E}}_{t+2}(y_{t+3})\big] \\
&\quad + \phi_y\phi_x\phi_x\mathbb{E}_{it}\big\{\bar{\mathbb{E}}_{t+1}\big[\mathbb{E}_{it+2}(x_{it+3})\big]\big\} + \phi_y\phi_x\phi_y\mathbb{E}_{it}\big\{\bar{\mathbb{E}}_{t+1}\big[\mathbb{E}_{it+2}(y_{t+3})\big]\big\} \qquad \text{(B.31)} \\
&\quad + \phi_y\phi_y\phi_x\mathbb{E}_{it}\big\{\bar{\mathbb{E}}_{t+1}\big[\bar{\mathbb{E}}_{t+2}(x_{it+3})\big]\big\} + \phi_y\phi_y\phi_y\mathbb{E}_{it}\big\{\bar{\mathbb{E}}_{t+1}\big[\bar{\mathbb{E}}_{t+2}(y_{t+3})\big]\big\} \\
&\quad + \phi_x\phi_x\mathbb{E}_{it}(\theta_{t+2}) + \phi_x\phi_y\mathbb{E}_{it}(\theta_{t+2}) + \phi_y\phi_x\mathbb{E}_{it}\big[\bar{\mathbb{E}}_{t+1}(\theta_{t+2})\big] + \phi_y\phi_y\mathbb{E}_{it}\big[\bar{\mathbb{E}}_{t+1}\big(\theta_{t+2}\big)\big] + (\phi_x+\phi_y)\mathbb{E}_{it}(\theta_{t+1}) + \theta_{it}.
\end{aligned}
\tag{B.32}
$$

Proceeding in such manner up to some arbitrary time J, one ends up getting a weighted sum of expectations of different orders about θ_{t+j}, for $j = 1, ..., J+1$, while the remaining endogenous variables x_{it+J+1} and y_{t+J+1} multiply coefficients that tend to zero as. Letting $J \to \infty$ and aggregating over agents, one finally obtains

$$
y_t = \sum_{j=1}^{\infty} \mathcal{P}^j_{\phi_x}\Big[(\phi_x+\phi_y)^j\bar{\mathbb{E}}^j(\theta_{t+j})\Big] \tag{B.33}
$$

where $\bar{\mathbb{E}}^j(\theta_{t+j})$ stands for the j^{th} order average expectation of θ_{t+j}, and, for notational convenience, we let $\bar{\mathbb{E}}^0(\theta_t) = \theta_t$. The way the operator $\mathcal{P}^j_{\phi_x}$ works is visible in the first four terms of the last line of (B.32) where $j = 2$. In the first and second term, ϕ_x appears as the first coefficient, which results in the expectation being of the first order. In the last two terms, ϕ_y appears as the first coefficient, which results in the expectation being of the second order. In subsequent substitution the pattern that results in the reduction of some of the higher order compounding is quite complex, as the combination of relative positions of ϕ_x and ϕ_y in the coefficients grows at the power of 2^n, so we omit it here. For instance, if one considers the expression at the 3rd iteration, i.e. for $J = 3$, then one has

$$
\begin{aligned}
y_t &= \big[\textbf{terms with } x_{it+4}\textbf{, } y_{t+4}\big] + \big(\phi_x\phi_x\phi_x + \phi_x\phi_x\phi_y\big)\bar{\mathbb{E}}_t(\theta_{t+3}) + +\big(\phi_x\phi_y\phi_x + \phi_x\phi_y\phi_y\big)\bar{\mathbb{E}}_t\big[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})\big] \\
&\quad + \big(\phi_y\phi_y\phi_y + \phi_y\phi_y\phi_x + \phi_y\phi_x\phi_y + \phi_y\phi_x\phi_x\big)\bar{\mathbb{E}}_t\big\{\bar{\mathbb{E}}_{t+1}\big[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})\big]\big\} \\
&\quad + \big(\phi_x\phi_x + \phi_x\phi_y\big)\bar{\mathbb{E}}_t(\theta_{t+2}) + \big(\phi_y\phi_x + \phi_y\phi_y\big)\bar{\mathbb{E}}_t\big[\bar{\mathbb{E}}_{t+1}\big(\theta_{t+2}\big)\big] + (\phi_x+\phi_y)\bar{\mathbb{E}}_t(\theta_{t+1}) + \theta_t.
\end{aligned}
\tag{B.34}
$$

Note that, together with the direct average expectation $\bar{\mathbb{E}}_t(\theta_{t+3})$, and the expectation of third order (the average expectation of the average expectation of the average expectation) $\bar{\mathbb{E}}_t\big\{\bar{\mathbb{E}}_{t+1}\big[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})\big]\big\}$, expression (B.34) also displays the average expectation at t of the average expectation at $t+2$, $\bar{\mathbb{E}}_t\big[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})\big]$. Substituting further, other combinations of higher order expectations compounding appear too. In summary, equation (B.34) shows that, in presence of dispersed information, the pattern of higher order expectations can be extremely cumbersome, and so the requirement of the canonical approach to work out each possible combination of expectations quickly becomes prohibitive.

B.5 Real Business Cycle Application: Full Model and Log-Linearization The model specification is taken from Graham and Wright (2010). The economy is structure in a continuum of islands indexed by $i \in [0, 1]$. Each island is inhabited by a representative household i and by a representative firm i. Household i supplies labor services exclusively to firm i in a decentralized competitive labor market. Labor of household i is the only labor productive in firm i. Households own capital in the economy and rent it out to firms in a centralized spot market. Capital, expressed in consumption goods, is productive in all the firms across the islands. The problem for Household i can be then written as

$$
\max_{C_{it}, K^{(s)}_{it+i}} \mathbb{E}_{it}\left\{\sum_{j=0}^{\infty}\beta^t\frac{C_{it}^{1-\eta}-1}{1-\eta}\right\}
$$

subject to a sequence of budget constraints of the form

$$C_{it} + K^{(s)}_{it+i} - (1-\delta) K^{(s)}_{it} = W_{it} + R_t K^{(s)}_{it},\ t = 0, 1, 2, ...,$$

where W_{it} is the wage rate in the labor market of island i and R_t is the rental rate of capital in the centralized capital market. Households are assumed, for the moment, to supply labor $N^{(s)}_{it}$ inelastically at the prevailing wage rate. We normalize the labor supplied by household i to $N^{(s)}_{it} = 1$. $K^{(s)}_{it+i}$ denotes the total capital that household i is bringing into period $t+1$. The superscript (s) stands for "supply" to denote the fact that the capital that household i is bringing into period $t+1$ will be the amount supplied by the same household in the centralized rental capital market at $t+1$. Symmetrically, in what follows the superscript (d) will stand for demand.

The problem for the representative firm in island i is

$$\max_{N^{(d)}_{it}, K^{(d)}_{it}} Y_{it} - W_{it} N^{(d)}_{it} - R_t K^{(d)}_{it}$$

where

$$Y^{(s)}_{it} = Z_{it} \left(K^{(d)}_{it}\right)^{\alpha} \left(N^{(d)}_{it}\right)^{1-\alpha}, \quad \alpha \in (0,1).$$

Output Y_{it} is supplied in the centralized market for output. In other words, in this economy there is only one consumption good centrally traded. The price of the consumption good at t is normalized to 1. Output is produced by firm i according to a Cobb-Douglas technology with labor and capital inputs and a technological factor Z_{it} that can be specified as

$$Z_{it} = e^{a_t + \varepsilon_{it}}.$$

The term a_t is common across all the islands, while ε_{it} is a productivity component that is specific to island i. The existence of a decentralized labor market together with an island specific productivity results in a labor income with an idiosyncratic risk component against which households would like to insure. We assume that markets for state contingent securities are not available, so that household i has to bear the labor income risk. We will also assume that the idiosyncratic labor income risk is not present in steady state, which means that the wealth distribution of the economy in steady state would be degenerate and an economy-wide representative household will exist. This is relevant at the linearization stage, as one would want to linearize the first order conditions for each island around the same steady state.

The first order conditions for firm i are

$$(1-\alpha) Y^{(s)}_{it} = N^{(d)}_{it} W_{it}$$

and

$$\alpha Y^{(s)}_{it} = K^{(d)}_{it} R_t,$$

so that the wage bill in island i is equal to a fraction $(1-\alpha)$ of output in the island, while the capital bill is the remaining fraction α. In addition, under the assumption that $N^{(s)}_{it} = 1$, market clearing for the decentralized labor market in island i, $N^{(s)}_{it} = N^{(d)}_{it}$, implies

$$Y^{(s)}_{it} = Z_{it} \left(K^{(d)}_{it}\right)^{\alpha} \text{ and } (1-\alpha) Y^{(s)}_{it} = W_{it}.$$

As a consequence, the participation of household i to the labor market, i.e. the observation of W_{it}, will result in the knowledge of $Y^{(s)}_{it}$. This will always be the case, independently of the equilibrium behavior of the variables.

The Euler equation for household i is

$$1 = \mathbb{E}_{it} \left[\beta \left(\frac{C_{it}}{C_{it+1}} \right)^{\eta} (R_t + (1-\delta)) \right],$$

while, using the first order conditions from firm i and the market clearing for the labor market, the budget constraint can be re-written as

$$C_{it} + K_{it+1}^{(s)} = (1-\alpha) Y_{it}^{(s)} + (R_t + (1-\delta)) K_{it}^{(s)}.$$

In addition, a no-Ponzi condition is assumed so that the solution path to the steady state has to satisfy the usual transversality condition.

To close the model in terms of market interactions we need to specify the market clearing condition for capital and for output, formally

$$\int_0^1 K_{it}^{(d)} di = \int_0^1 K_{it}^{(s)} di$$

and

$$\int_0^1 C_{it} di + \int_0^1 \left(K_{it+1}^{(s)} - (1-\delta) K_{it}^{(s)} \right) di = \int_0^1 Y_{it}^{(s)} di.$$

We will work with a log-linearized version of the economy around a steady state that is derived under the assumption that the long run unconditional average of Z_{it} is 1. Notice that this implies that the economy does not display a growth trend in steady state. This is without loss of generality for the purpose of the application.

For any variable X_t we define x_t as $X_t = X^* e^{x_t}$, where X^* is the steady state value of X_t. The log-linearized economy is given by the following set of equations (details available from the authors upon request). The budget constraint for household i is

$$\left(\frac{1}{\alpha} (\rho + \delta) - \delta \right) c_{it} + k_{it+1}^{(s)} = \frac{1-\alpha}{\alpha} (\rho + \delta) y_{it}^{(s)} + (\rho + \delta) r_t + \frac{1}{\beta} k_{it}^{(s)} \tag{B.35}$$

where $\rho = \frac{1}{\beta} - 1$ is the rate of time preference. The Euler equation for household i is

$$\mathbb{E}_{it} \left[\eta (c_{it} - c_{it+1}) + \beta (\rho + \delta) r_{t+1} \right] = 0. \tag{B.36}$$

Output supplied by firm i is given by

$$y_{it}^{(s)} = a_t + \varepsilon_{it} + \alpha k_{it}^{(d)}, \tag{B.37}$$

and the rental rate for capital is

$$r_t = a_t + \varepsilon_{it} + (\alpha - 1) k_{it}^{(d)}. \tag{B.38}$$

The demand for capital of firm i is

$$k_{it}^{(d)} = w_{it} - r_t^K \tag{B.39}$$

and the wage rate is

$$w_{it} = y_{it}^{(s)}. \tag{B.40}$$

The market clearing condition for aggregate capital is

$$\int_0^1 k_{it}^{(d)} di = \int_0^1 k_{it}^{(s)} di = k_t \tag{B.41}$$

while the market clearing for aggregate output is

$$\left(\frac{1}{\alpha} (\rho + \delta) - \delta \right) c_t + (k_{t+1} - (1-\delta) k_t) = \frac{1}{\alpha} (\rho + \delta) y_t \tag{B.42}$$

where

$$c_t = \int_0^1 c_{it} di, \;\; y_t = \int_0^1 y_{it}^{(s)} di.$$

The set of equations (B.35)-(B.42) completely describe the equilibrium dynamics of the linearized economy, conditional on the sequence of cross sectional distributions of information sets implicit in the expectational operator $\mathbb{E}_{it}$ for $i \in [0,1]$ and $\forall t$. The equations presented in the text are a direct application of the equations above with $\eta = 1$, $\delta = 1$, and the notation $h_{it} = k_{it}^d$.

C Online Technical Appendix for Rondina and Walker (2016)

Elementary results concerning the theory of stationary stochastic processes and the residue calculus are necessary for grasping the z-transform approach advocated in Rondina and Walker (2016). The purpose of this appendix is to offer readers unfamiliar with the methods used in Rondina and Walker (2016) the additional background necessary such that the paper is self-contained. The appendix introduces important theorems that are relatively well known but is by no means exhaustive. Interested readers are directed to Brown and Churchill (2013) and Whittle (1983) for good references on complex analysis and stochastic processes. Sargent (1987) provides a good introduction to these concepts and discusses economic applications.

C.1 Variance-Covariance Generating Function Consider two-covariance stationary linear-Gaussian multivariate processes, $\{\omega_t, t \in \mathbb{Z}\}$ and $\{s_t, t \in \mathbb{Z}\}$, where the vector dimensions are $n \times 1$, and $m \times 1$, respectively. Let $\Upsilon_{\omega s}(j)$ denote the $m \times n$ unconditional covariance matrix between ω_t and s_{t-j}, for $j \in \mathbb{Z}$, formally

$$\Upsilon_{\omega s}(j) \equiv \mathbb{E}(\omega_t s_{t-j}^T) - \mathbb{E}(\omega_t)\mathbb{E}(s_{t-j}^T), \tag{C.1}$$

where T denotes transpose. The variance-covariance generating function is then defined as

$$g_{\omega s}(z) \equiv \sum_{j=-\infty}^{\infty} \Upsilon_{\omega s}(j) z^j, \tag{C.2}$$

where $g_{\omega s}(z)$ is an $m \times n$ matrix. When $\omega_t = s_t$ the function is referred to as the auto-covariance generating function and denoted by $g_{\omega\omega}(z)$, or $g_{ss}(z)$. An extensive treatment of the properties of the variance-covariance generating function can be found in Sargent (1987).

C.2 Wold Fundamental Representation Theorem Much of the analysis in Rondina and Walker (2016) is conducted in the space of lag polynomials without specific functional forms assumed (e.g., $ARMA(m, n)$). The Wold representation theorem allows for such a general specification.

Theorem C1. *[Wold Representation Theorem] Let $\{s_t\}$ be any $(n \times 1)$ covariance stationary stochastic process with $\mathbb{E}(s_t) = 0$. Then it can be uniquely represented in the form*

$$s_t = \eta_t + \Gamma^*(L)\hat{w}_t \tag{C.3}$$

where $\Gamma^*(L)$ is a matrix polynomial in the lag operator, and $\sum_{j=0}^{\infty} \Gamma_j^* \Gamma_j^{*T}$ is convergent. The process $\hat{w}_t$ is n-variate white noise with $\mathbb{E}(\hat{w}_t) = 0$, $\mathbb{E}(\hat{w}_t \hat{w}_t') = I_n$ and $\mathbb{E}(\hat{w}_t \hat{w}_{t-m}') = 0$ for $m \neq 0$. The process $\Gamma_0^* \hat{w}_t$ is the innovation in predicting s_t linearly from its own past:

$$\Gamma_0^* \hat{w}_t = s_t - P(s_t | s_{t-1}, s_{t-2}, ...), \tag{C.4}$$

where $P(\cdot)$ denotes linear projection. The process η_t is linearly deterministic; there exists an $n \times 1$ vector c_0 and $n \times n$ matrices C_s such that without error $\eta_t = c_0 + \sum_{s=1}^{\infty} C_s \eta_{t-s}$ and $\mathbb{E}[\hat{w}_t \eta_{t-m}'] = 0$ for all m.

The Wold representation theorem states that *any* covariance stationary process can be written as a linear combination of a (possibly infinite) moving average representation where the innovations are the linear forecast errors for s_t and a process that can be predicted arbitrarily well by a linear function of past values of s_t. The theorem is a *representation* determined by second moments of the stochastic process only and therefore may not fully capture the data generating process. For example, that the decomposition is linear suggests that a process could be deterministic in the strict sense and yet linearly non-deterministic; Whittle (1983) provides examples of such processes. The innovations in the Wold representation are generated by linear projections which need not be the same as the conditional expectation ($\mathbb{E}[s_t | s_{t-1}, s_{t-2}, ...]$). However, when working with linear Gaussian stochastic processes, as is standard in the rational expectations literature, the best conditional expectation coincides with linear projection.

As discussed in detail in the paper, the innovations derived from the Wold representation are an essential element of a rational expectations equilibrium in that they define the information set obtained from conditioning on current and past s_t's. Using the language of Rozanov (1967), the innovation sequence $\{\hat{w}_{t-j}\}_{j=0}^{\infty}$ of (C.4) are fundamental for the sequence $\{s_{t-j}\}_{j=0}^{\infty}$ if the Hilbert space generated by the observables is equivalent (in mean-square norm) to the Hilbert space generated by the innovations.

C.3 Canonical Factorization Let s_t be specified as

$$s_t = \Gamma(L) u_t, \tag{C.5}$$

where $\Gamma(L)$ is a lag polynomial matrix with square-summable coefficients, and u_t is a $m \times 1$ vector of i.i.d. Gaussian innovations with zero mean and variance-covariance matrix Σ_u. Testing for fundamentalness in a process like (C.5) can be done by checking for the invertibility of $\Gamma(z)$ for $|z| < 1$.[19] While testing for fundamentalness is straightforward, deriving the unique fundamental Wold representation is not. Here we rely on powerful factorization theorems. The following definition is taken from Whittle (1983). If $g(z)$ is a square matrix function of z, then the canonical factorization

$$g(z) = \Gamma^*(z)\Gamma^*(z^{-1})^T, \tag{C.6}$$

will refer to a factorization in which both $\Gamma^*(z)$ and its inverse have valid expansions in non-negative powers of z for $|z| \leq 1$; and $\Gamma^*(z^{-1})^T$ and its inverse have expansions in non-positive powers of z for $|z| \geq 1$. The canonical factorization of the variance-covariance generating function delivers the Wold fundamental representation of the time series. It yields both the fundamental moving average Wold representation, and the autoregressive representation that is consistent with span $\{s_{t-j}\}_{j=0}^{\infty}$ being equivalent to span $\{\hat{w}_{t-j}\}_{j=0}^{\infty}$.

There does not exist a general method to compute analytically the canonical factorization of any arbitrary matrix $g(z)$. Several methods that have been proposed work well when the knowledge of the structure of the matrix $g(z)$ is used. Rozanov (1967), page 47, proposes an algorithm to obtain $\Gamma^*(z)$ from $\Gamma(L)$ that can be used when $\Gamma(L)$ is known to have a finite number of isolated singularities $\lambda_1, \lambda_2, ..., \lambda_p$ inside the unit circle. In the case of $n = 2$, and one singularity at $|\lambda| < 1$, one has that

$$\Gamma^*(z) = \Gamma(z) W_\lambda B_\lambda(z), \tag{C.7}$$

where $B_\lambda(z)$ is the Blaschke matrix

$$B_\lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1-z\lambda}{z-\lambda} \end{bmatrix}. \tag{C.8}$$

The constant matrix W_λ is a unitary matrix, so that $W_\lambda^{-1} = W_\lambda^T$, whose columns are the left singular vectors of the system matrix $\Gamma(z)$ evaluated at λ. While the inverse of $\Gamma(\lambda)$ does not exist, the pseudo-inverse (also known as the Moore-Penrose inverse) always exists and it is given by

$$\Gamma(\lambda)^+ = U\Sigma V^T, \tag{C.9}$$

where $V\Sigma U^T$ is the singular value decomposition of the matrix $\Gamma(\lambda)$. The left singular vectors of $\Gamma(\lambda)$ are the columns of U. The following result, which can be showed by using widely available formulas for the singular value decomposition of 2 by 2 matrices is used in the proof of Lemma A1.

Lemma C1. *Consider the matrix*

$$N = \begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix}. \tag{C.10}$$

Then the matrix U in the singular value decomposition of $N = V\Sigma U^T$, is

$$U = \frac{1}{\sqrt{c^2+d^2}} \begin{bmatrix} \frac{cd}{|d|} & -\frac{cd}{|c|} \\ \frac{d^2}{|d|} & \frac{c^2}{|c|} \end{bmatrix}. \tag{C.11}$$

[19] There are many ways to test for fundamentalness, see Fernandez-Villaverde, Rubio, Sargent, and Watson (2007).

C.4 Riesz-Fischer Theorem

Theorem C2. *[Riesz-Fischer] Let $D(\sqrt{r})$ denote a disk in the complex plane of radius $\sqrt{r}$ centered at the origin. There is an equivalence (i.e. an isometric isomorphism) between the space of r-summable sequences $\sum_j r^j |f_j|^2 < \infty$ and the Hardy space of analytic functions $f(z)$ in $D(\sqrt{r})$ satisfying the restriction*

$$\frac{1}{2\pi i}\oint f(z)f(rz^{-1})\frac{dz}{z} < \infty$$

where $\oint$ denotes (counterclockwise) contour integration around $D(\sqrt{r})$. An analytic function satisfying the above condition is said to be r-integrable.[20]

The Riesz-Fischer theorem ensures that one can work either in the space of infinite sequence of square-summable matrix coefficients $\{..., Q_{-1}, Q_0, Q_1, ...\}$, or in the space of complex-valued analytic functions $Q(z) = \sum_{j=-\infty}^{\infty} Q_j z^j$, since the two spaces are equivalent.

C.5 Wiener-Kolmogorov Prediction Formula

Consider the problem of computing the linear least-squares estimate for ω_{t+j}, denoted by $\hat{\omega}_{t+j}$, conditional on the realized history $\{s_{t+j}\}_{j=0}^{\infty}$. The solution to the problem consists in the sequence of real valued matrices $\{\Pi_j\}_{j=0}^{\infty}$, or, equivalently, the complex-valued function $\Pi(z)$, such that

$$\hat{\omega}_{t+j} = \sum_{j=0}^{\infty} \Pi_j s_{t-j}. \tag{C.12}$$

We assume that all the processes have zero unconditional mean. The following result is a version of the Wiener-Kolmogorov prediction formula taken from Whittle (1983).

Theorem C3. *Suppose that $g_{ss}(z)$ has the canonical factorization*

$$g_{ss}(z) = \Gamma^*(z)\Gamma^*(z^{-1})^T. \tag{C.13}$$

then the generating function $\Pi_j(z) = \sum_{i=0}^{\infty} \Pi_i z^i$ of the optimal estimate $\hat{\omega}_{t+j}$ is

$$\Pi_j(z) = \left[z^{-j} g_{\omega s}(z)\left(\Gamma^*(z^{-1})^T\right)^{-1}\right]_+ \Gamma^*(z)^{-1}. \tag{C.14}$$

Proof. The minimization of the the means squared forecast errors $\mathbb{E}(\hat{\omega}_t - \omega_t)^2$ leads to the set of first order conditions

$$\sum_{k=0}^{\infty} \Upsilon_{ss}(i-k) = \Upsilon_{\omega s}(i-j), \quad \text{for} \quad i = 0, 1, 2, ... \tag{C.15}$$

Multiplying both sides by z^i and adding over all integral i's one gets the Wiener-Hopf relationship

$$\Pi_j(z) g_{ss}(z) = z^{-j} g_{\omega s}(z) + h(z), \tag{C.16}$$

where $h(z) = \sum_{i=-\infty}^{-1} h_i z^i$, is an unknown matrix series in negative powers of z. If we post-multiply both sides by $\left[\Gamma^*(z^{-1})^T\right]^{-1}$, we obtain

$$\Pi_j(z)\Gamma^*(z) = z^{-j} g_{\omega s}(z)\left[\Gamma^*(z^{-1})^T\right]^{-1} + h(z)\left[\Gamma^*(z^{-1})^T\right]^{-1}, \tag{C.17}$$

Note that both sides can be represented as matrix series in powers of z whose coefficients have to obey the above relationship. By construction, the left hand side has only positive powers of z, while the second term of the right hand side has only negative powers of z. If we apply the annihilating operator on both sides we thus get

$$\Pi_j(z)\Gamma^*(z) = \left[z^{-j} g_{\omega s}(z)\left[\Gamma^*(z^{-1})^T\right]^{-1}\right]_+. \tag{C.18}$$

[20] This theorem is usually proved for the case $r = 1$ and for functions defined on the boundary of a disk. For further exposition see Sargent (1987).

The last step consists in post-multiplying both sides by $\Gamma^*(z)^{-1}$ so that (C.14) obtains. □

The requirement that $\Gamma^*(z)$ should be the canonical factorization of $g_{ss}(z)$ is essential in two steps. First, from it we ensure that $\left[\Gamma^*(z^{-1})^T\right]^{-1}$ has an expansion in non-positive powers of z, which means that the term $h(z)\left[\Gamma^*(z^{-1})^T\right]^{-1}$ disappears when the annihilating operator is applied. Second, we also ensure that $\Gamma^*(z)^{-1}$ has an expansion in non-negative powers of z, which result in $\Pi(z)$ having an expansion in non-negative powers of z only, as required.

C.6 Whiteman (1983) Solution Method: Existence and Uniqueness of REE We use the existence and uniqueness criteria of Whiteman (1983) developed for linear, rational expectations equilibria. The following works through the three relevant cases–existence but no uniqueness, no existence, and existence-uniqueness. Consider the following generic rational expectations model

$$\zeta\tilde{\zeta}\mathbb{E}_t y_{t+1} - (\zeta + \tilde{\zeta})y_t + y_{t-1} = \theta_t, \qquad \theta_t = A(L)\varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0,1) \tag{C.19}$$

where ε_t is assumed to be fundamental for θ_t (i.e., $A(L)$ is assumed to have a one-sided inverse in non-negative powers of L). Following the solution principle, we will look for a solution that is square-summable in the Hilbert space generated by the fundamental shock ε, $y_t = Q(L)\varepsilon_t$ (third tenet). If we invoke the optimal prediction formula (C.14), then $\mathbb{E}_t y_{t+1} = [Q(L)/L]_+\varepsilon_t = L^{-1}[Q(L) - Q_0]\varepsilon_t$. Together with the fourth tenet of the solution principle (i.e., that the rational expectation restrictions hold for all realizations of ε), this implies that (C.19) can be written in z-transform as

$$z^{-1}[Q(z) - Q_0]\zeta\tilde{\zeta} - (\zeta + \tilde{\zeta})Q(z) + zQ(z) = A(z)$$

Multiplying by z and rearranging delivers

$$Q(z) = \frac{zA(z) + Q_0}{(\zeta - z)(\tilde{\zeta} - z)} \tag{C.20}$$

Appealing to the Riesz-Fischer Thereom, square-summability (stationarity) in the time domain is tantamount to analyticity of $Q(z)$ on the unit disk. The function $Q(z)$ is analytic at z_0 if it is continuously (complex) differentiable in an open neighborhood of z_0.[21] Any rational function $(f(z)/g(z))$ where $f(\cdot)$ and $g(\cdot)$ are polynomials will be analytic on the unit disk provided $g(z) \neq 0$ at any point inside the unit circle. The extent to which this is true for $Q(z)$ depends upon the parameters ζ and $\tilde{\zeta}$.

As shown in Whiteman (1983), there are three cases one must consider. First, assume that $|\zeta| > 1$ and $|\tilde{\zeta}| > 1$. Then (C.20) is an analytic function on $|z| < 1$ and the representation is given by

$$y_t = \left[\frac{LA(L) + Q_0}{(\zeta - L)(\tilde{\zeta} - L)}\right]\varepsilon_t \tag{C.21}$$

For any finite value of Q_0, this is a solution that lies in the Hilbert space generated by $\{\theta_t\}$ and satisfies the tenets of the solution principle. Thus, we have existence but not uniqueness because Q_0 can be set arbitrarily.

The second case to consider is $|\tilde{\zeta}| < 1 < |\zeta|$. In this case, the function $Q(z)$ has an isolated singularity at ζ, implying that $Q(z)$ is not analytic on the unit disk. In this case, the free parameter Q_0 can be set to remove the singularity at ζ by setting Q_0 in such a way as to cause the residue of $Q(\cdot)$ to be zero at ζ

$$\lim_{z\to\zeta}(\tilde{\zeta} - z)Q(z) = \frac{\zeta A(\zeta) + Q_0}{\zeta - \tilde{\zeta}} = 0$$

Solving for Q_0 delivers $Q_0 = -\zeta A(\zeta)$. Substituting this into (C.21) yields the following rational expectations equi-

[21] Analytic is synonymous with holomorphic, regular and regular analytic.

librium

$$y_t = \left[\frac{LA(L) - \zeta A(\zeta)}{(\zeta - L)(\tilde{\zeta} - L)} \right] \varepsilon_t \tag{C.22}$$

The function $Q(z)$ is now analytic and (C.22) is the unique solution that lies in the Hilbert space generated by $\{\theta_t\}$. The solution is the ubiquitous Hansen-Sargent prediction formula that clearly captures the cross-equation restrictions that are the "hallmark of rational expectations models" [Hansen and Sargent (1980)].[22]

The final case to consider is $|\zeta| < 1$ and $|\tilde{\zeta}| < 1$. In this case, (C.20) has two isolated singularities at ζ and $\tilde{\zeta}$, and Q_0 cannot be set to remove both singularities.[23] Hence in this case, there is no solution in the Hilbert space generated by $\{\theta_t\}$ and we do not have existence.

C.7 Annihilating Operator Let $H(z) = \sum_{m=-\infty}^{+\infty} H_m z^m$. The following Lemma is due to Hansen and Sargent (1980).

Lemma C2. *Let $H(z)$ be a regular function in $|z| < 1$, with at most a finite number of singularities $z_1, z_2, ..., z_k$ in $|z| < 1$. Let $\pi_1(z)$, $\pi_2(z)$,..., $\pi_k(z)$ denote the principal parts of the Laurent series expansion of $H(z)$ around the singularities. Then*

$$H(z) - \sum_{j=1}^{k} \pi_j(z). \tag{C.23}$$

The Laurent series expansion of $H(z)$ around the singularity z_j is

$$H(z) = \sum_{m=-\infty}^{+\infty} H_m (z - z_j)^m, \tag{C.24}$$

and its principal part is given by

$$\pi_j(z) = \sum_{-\infty}^{-1} H_m (z - z_j)^m. \tag{C.25}$$

To compute the principal part at z_j first compute the constant $\hat{\pi}_j$ as

$$\hat{\pi}_j = \lim_{z \to z_j} (z - z_j) H(z), \tag{C.26}$$

and then set

$$\pi_j(z) = \hat{\pi}_j (z - z_j). \tag{C.27}$$

[22] Our methodology can also handle unit roots. For example, suppose $\theta_t = (1 - L)A(L)\varepsilon_t$. The solution, $Q(L)\varepsilon_t$, would then inherit the unit root via the cross-equation restriction.

[23] As discussed by Whiteman (1983), the problem remains even if $\zeta = \tilde{\zeta}$.